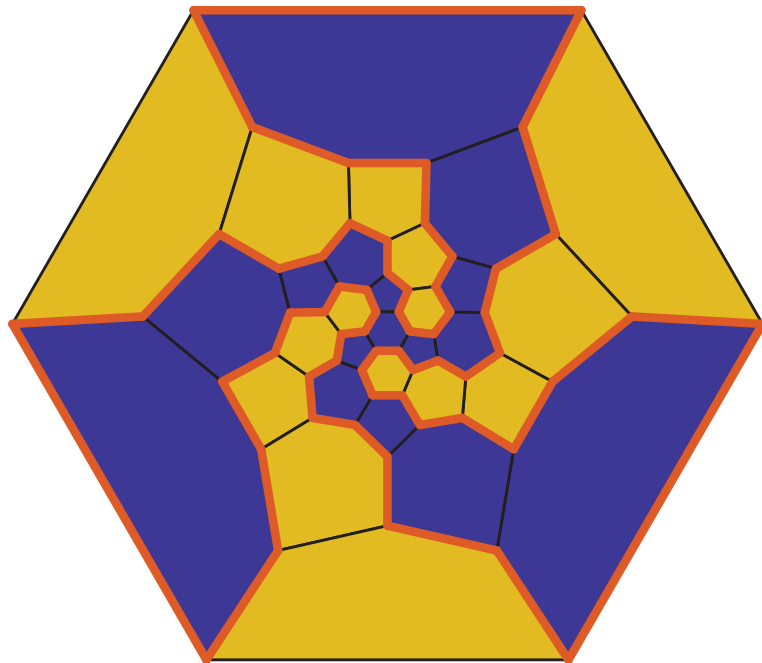


Games, Graphs, and Groups

An Introduction



Wednesday December 24, 2025

Matthias Weber

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Contents

1	Combinatorial Game Theory	3
1.1	Introduction	3
1.2	Tutorial: Number Systems	6
1.3	The Game NIM	11
1.4	NIM in Disguise	18
1.5	Sprague-Grundy Theory	26
1.6	More Examples of Impartial Games	31
1.7	Coin-Turning Games	40
2	Polyforms	42
2.1	Polyominoes	43
2.2	Polyiamonds	58
2.3	Polyhexes	69
3	Dissection Puzzles and Periodic Tilings	75
3.1	Paradoxical Dissections	75
3.2	A Taste of Linear Algebra	78
3.3	Archimedean Tilings	84
3.4	Which Shapes Tile?	95
3.5	Dissections	99
3.6	Polyform Dissections	110
3.7	Napoleon's Theorem	115
3.8	Morley's Theorem	118
4	Playing with Graphs and Maps	121
4.1	Graphs	121
4.2	River Crossings	132
4.3	Euler Circuits	136
4.4	Maps and Euler's Theorem	139
4.5	Coloring Graphs and Maps	146
4.6	Brussel Sprouts	150
5	Solitaire Games	151
5.1	Peg Solitaire	151
5.2	Permutations	154
5.3	The 14-15 Puzzle	158
5.4	Conway's Army	160
5.5	Instant Insanity	164
6	Appendix	167
6.1	Bibliography	167
6.2	Index	168

Introduction

*"Need must find," said the sea-thing child.
"Find what?" said the crab.
"Whatever there is," said the sea-thing child.
RUSSEL HOBAN, *The Sea-Thing Child**

In a physical world where most geographic discoveries have been made, exploration has become an occupation of the mind. In the sciences, this is called research, and, due to the effort of innumerable fearless scientists, the frontiers of scientific exploration have been pushed beyond the reach of young aspiring explorers.

This is in particular true for the mathematical sciences, because Mathematics is largely vertical: Taking a new step requires that you have mastered all the other steps below. Theorems are based on other theorems in endless chains.

So, how does one get an idea of what research is like without undergoing many years of graduate studies? The answer is to walk off the trodden path, in an area that is too much fun to be taken seriously by the professional explorers: The area of mathematical puzzles and games.

There are many fascinating books about mathematical puzzles and games, and this text borrows from many of them. The idea was to select those games and puzzles that are amenable to a mild mathematical treatment, allow for exercises that practice whatever skills are necessary to master these, and, most importantly, are fertile enough so that the creative mind is able to stake out new territory.

The result is a handful of loosely connected chapters, each dedicated to a circle of ideas. If necessary, there are sections about background material like graphs, periodic tilings, or the number system. Solutions and strategies for puzzles and games are used to introduce concepts that are then applied and practiced in exercises. Finally, there are activities and explorations in each chapter: Activities are things you can do ad hoc, while explorations are open-ended exercises where the learned methods promise some progress towards a solution, but not necessarily a complete solution. I have tried hard to come up with explorations for which I do not know the answer. Hence, each chapter attempts to mimic the three steps of any expedition: the acquisition of tools, the practice of skills, and the leap into the unknown.

I have been writing this text since 2001 to accompany the class *Exploring Mathematical Ideas* at Indiana University, Bloomington. This used to be an Intensive Writing class, and the writing projects for the students essentially consist of expedition diaries for the various explorations they get assigned. The hardest part is usually convincing the students that, for the quality of the essays, what matters most is the spirit, rather than the results. Coming up with five new ideas that all fail shows a better spirit than a solution that applies a well-established method.

The chapters in this text are largely independent. For teaching purposes, I found that beginning with games helps to bring the students together. After all, collaboration is a key ingredient to any successful expedition. Besides that, the chapter about periodic tilings uses the earlier introduced polyforms as examples, and *Playing with Graphs* uses some basic ideas about strategies that are explained in the chapter about impartial games.

The text comes with many exercises, and I have sometimes included solutions.

Bloomington, Autumn 2025

1 Combinatorial Game Theory

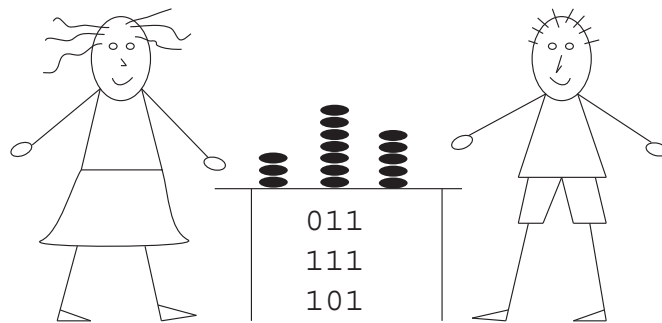


Figure 1.1

1.1 INTRODUCTION

Humans enjoy playing various games, including baseball, poker, Quidditch, chess, and many others. While Mathematics has to say *something* about most games, those games we actually play are (fortunately) too complicated to be completely solvable by Mathematics. Therefore, we will begin by studying some rather simple games to see how Mathematics can help.

Let's list the restrictions we'd like to impose:

Definition 1.1.1 A combinatorial game satisfies the following:

1. The game will be for two players only. This eliminates soccer, for instance.
2. The game is sequential, i.e., the players take turns and can't move simultaneously. This eliminates many games of skill.
3. The game has perfect information. This means that both players know at each move the complete state of the game, i.e., they are both able to determine all possible moves. This eliminates most card games that involve chance.

While very restrictive, this still leaves us with games like chess or Go, and in fact, most strategy games.

If we think about such combinatorial games, what really counts are the possible positions the game can have, and, for each player and each position, the possible moves the player has. These possible moves to new positions are called *options*.

Definition 1.1.2 A combinatorial 2-person game is given by a

1. a set of positions;
2. a distinguished start position;
3. for each position and each player a (possibly empty) set of options, i.e., positions the player can move to in one legal move;
4. for each ending position, i.e., a position without options, a choice whether this is a losing or winning position for the first player.

The gameplay then proceeds with the two players taking turns, beginning at the starting position, and each choosing from their set of options. A turn is also called a *move*. The game ends (usually) with the player losing who has no options left, or when the game reaches a position that has been designated as a losing position.

finite
game

Definition 1.1.3 We call a game *finite* if any sequence of legal moves has finite length. Finite games necessarily stop.

From these definitions, it is clear that a finite game will end with a win of either player — there are no draws.

strategy

Definition 1.1.4 A strategy for a player is a choice of an option for that player in each game position. A winning strategy is a strategy that always ends with a win for that player. A position is a winning position if there is a winning strategy for the first player. All other positions are losing positions.

These concepts are best illustrated using a *game tree*, which we will introduce by means of the following simple game:

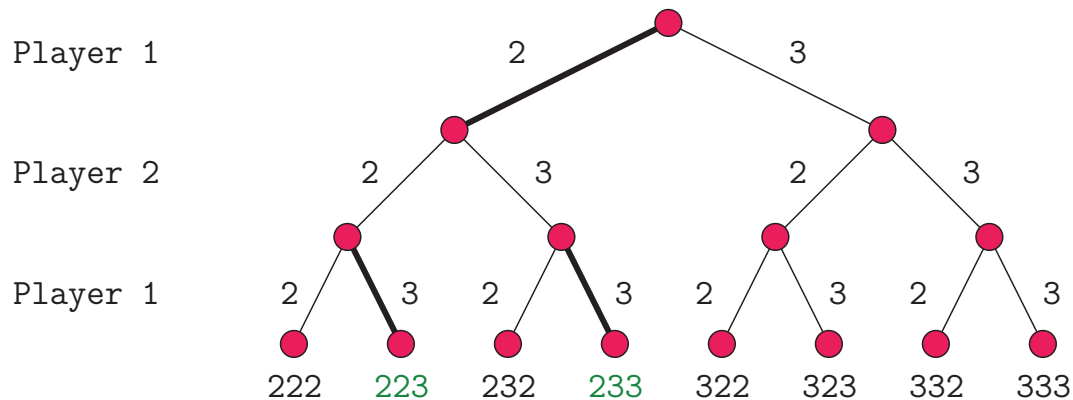


Figure 1.2 A simple game tree

Example 1.1.1 Two players create a 3-digit number by choosing the digit 2 or 3 and taking turns writing the number down from left to right. The first player wins if the final number is a prime number. At first, the chances for the first player look bleak, as only two of the eight final numbers are prime numbers, namely 223 and 233. But the winning strategy consists of choosing the fat edges in the game tree of **Figure 1.2** as moves. All other positions are irrelevant. A winning strategy needs to provide answers for all countermoves the opponent can offer. The first move — the winning move — picks the digit 2, choosing the first branch of the tree. Then the second player picks either a left or a right branch, and in each case, there is a response for the first player that leads to a prime number.

Exploration 1.1.2 Create and analyze variations of this game. For instance:

1. What happens with trees that are deeper (lead to 4-digit numbers, for instance)?
2. What happens if you pick other digits for the tree?
3. What happens if you allow choosing from more than two digits and every move?
4. Instead of winning at a prime number, think of other winning conditions.

Theorem 1.1.5 *In a finite two-person game, either the first or the second player has a winning strategy.*

Proof. Let's assume that the second player does *not* have a winning strategy. We have to find a winning strategy for the first player. The first player chooses a move so that the second player has no winning strategy from the resulting position. Such a move must exist, because otherwise the second player would have a winning strategy. Then we consider all options for the second player, and repeat the argument. As the game play ends, we will have assigned a winning move to any position that occurs in this game play by following the strategy of the first player. For the remaining positions that won't actually occur, we can assign any move we like to the first player. $\square$

We can illustrate this proof using a game tree, too: For a finite game, draw the game tree (which is usually very, very large). Mark all ending positions as losing or winning positions. Then, proceed by marking all vertices of the tree from which you can reach one of the already marked losing positions by following an edge, i.e., by making a move, as a winning position. In contrast, if you can reach from a vertex only winning positions, mark that vertex as a losing position (because the player who has to move must move into a position which is a winning position for the next player). Repeat this process until the position you are interested in has been marked. This must happen eventually, as the game is assumed to be finite.

This means we have an algorithm that will eventually decide which player has a winning strategy for any game position. There are two problems with this, however: Firstly, this algorithm is very time-consuming and impractical, even for relatively simple games. Secondly, it only deals with individual games. Mathematicians like to make statements about entire classes of games. We will do so by studying *impartial games*. Before that, we will spend a little time reviewing number systems. If you are familiar with that, you can skip to the next ??sec:nim.

1.2 TUTORIAL: NUMBER SYSTEMS

When we write an integer like 6732, this is just shorthand for

$$6732 = 6 \cdot 10^3 + 7 \cdot 10^2 + 3 \cdot 10^1 + 2 \cdot 10^0$$

which makes explicit that we are using the decimal number system. That we are using the number 10 as a *base* is just a remnant from the elementary counting process using our ten fingers. Any other positive integer can serve as a base as well. For instance, if we had just two fingers, we might have chosen the base for our number system to be 2, and we had only two digits, 0 and 1. The *binary number* 11001_2 would then mean

$$11001_2 = 1 \cdot 2^4 + 1 \cdot 2^3 + 0 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

Note that we have started indicating the base 2 by writing it as an index, to avoid confusing the binary number 11001_2 with the decimal number 11001_{10} . We will only indicate the decimal base for emphasis. Converting a number in binary form to decimal form is easy as long as we remember a few powers of two:

n	0	1	2	3	4	5	6	7	8	9	10
2^n	1	2	4	8	16	32	64	128	256	512	1024

Table 1.1 The powers of 2

Exercise 1.2.1 Convert the following binary numbers into decimal form:

1. 101_2
2. 1111111_2
3. 1010101_2

To convert a decimal number into its binary form, we can either proceed from the left to the right, or from the right to the left. We will learn both ways, and you can use the method you like best.

We will first discuss how to find the binary form of 1234, producing the digits from left to right. This means we have to find the binary digits $b_0, b_1, \dots$ such that

$$1234_{10} = \dots + b_4 \cdot 2^4 + b_3 \cdot 2^3 + b_2 \cdot 2^2 + b_1 \cdot 2^1 + b_0 \cdot 2^0$$

In other words, we have to write 1234 as a sum of powers of 2. To this end, we find the highest power of 2 that fits into 1234, which is $2^{10} = 1024$ (we can use the table for that). So we found that the first digit is $b_{10} = 1$. This leaves us to write $210 = 1234 - 1024$ as a sum of powers of 2. Again, the highest power of 2 fitting into 210 is $2^7 = 128$. This tells us that $b_7 = 1$ and $b_8 = b_9 = 0$, because we didn't use 2^8 or 2^9 . We keep going like this:

$$1234 = 1024 + 210 = 2^{10} + 210$$

$$210 = 128 + 82 = 2^7 + 82$$

$$82 = 64 + 18 = 2^6 + 18$$

$$18 = 16 + 2 = 2^4 + 2$$

$$2 = 2 + 0 = 2^1$$

so that

$$\begin{aligned} 1234_{10} &= 2^{10} + 2^7 + 2^6 + 2^4 + 2^1 \\ &= 10011010010_2 \end{aligned}$$

Make sure to fill in all zeroes!

Exercise 1.2.2 *Convert the following decimal numbers into binary form:*

1. 111_{10}
2. 222_{10}
3. 333_{10}
4. 999_{10}
5. 1023_{10}

This method is relatively quick if you are familiar with the powers of two and have a good grasp of subtraction.

We now turn to the second method that produces the binary digits from right to left. For motivation, let's look again at the decimal number 6732. Dividing this number by 10 gives the quotient 673 with a remainder of 2. So the last digit 2 is just the remainder after division by 10. The second last digit 3 can be found by dividing the quotient 673 again by 10, and taking its remainder. So, by rewriting

$$\begin{aligned} 6732 &= 10 \cdot 673 + 2 \\ &= 10 \cdot (10 \cdot 67 + 3) + 2 \\ &= 10 \cdot (10 \cdot (10 \cdot 6 + 7) + 3) + 2 \end{aligned}$$

we can recover all digits of 6732 as the remainders after successive divisions by 10. Similarly, in the binary system, writing

$$11001_2 = 2 \cdot (2 \cdot (2 \cdot (2 \cdot 1 + 1) + 0) + 0) + 1$$

shows that we can also recover the binary digits of a number as the remainders after successive divisions by 2. Let's use this to convert again 1234 into its binary form:

$$\begin{array}{ll} 1234 \div 2 = 617 & \text{remainder } 0 \\ 617 \div 2 = 308 & \text{remainder } 1 \\ 308 \div 2 = 154 & \text{remainder } 0 \\ 154 \div 2 = 77 & \text{remainder } 0 \\ 77 \div 2 = 38 & \text{remainder } 1 \\ 38 \div 2 = 19 & \text{remainder } 0 \\ 19 \div 2 = 9 & \text{remainder } 1 \\ 9 \div 2 = 4 & \text{remainder } 1 \\ 4 \div 2 = 2 & \text{remainder } 0 \\ 2 \div 2 = 1 & \text{remainder } 0 \\ 1 \div 2 = 0 & \text{remainder } 1 \end{array}$$

which gives the same answer 10011010010_2 as before. Don't forget that the digits are found from the right to the left! This method is great if you are good at dividing by two. It also has the advantage that you don't need to know the powers of two, so it will be better for larger numbers. Let's practice!

Exercise 1.2.3 Convert the following decimal numbers into binary form, using the division method:

1. 1111_{10}
2. 2222_{10}
3. 3333_{10}
4. 9999_{10}
5. 123456789_{10}

The same methods work to convert numbers between any number systems. We begin with a few straightforward exercises and then move to some trickier problems. For number systems that have a base b larger than 10, we need digits to designate the numbers $10, 11, \dots, b - 1$. It is customary to use the letters $A, B, C, \dots$ for that. So, for example, $2B_{16} = 2 \cdot 16 + 11 = 43$.

Exercise 1.2.4 Write as decimal numbers:

1. 121212_3
2. 123123_4
3. ABC_{16}

Exercise 1.2.5 Write the decimal number 1234567 in the number systems with base 3, 7, and 13.

Exercise 1.2.6 ([1]) A visitor from Mars overhears a school teacher demonstrating that the equation

$$5x^2 - 50x + 125$$

has only the solution $x = 5$. They protest that there would be a second solution. How many fingers do Martians have?

Solution: This looks like an impossible problem: Surely, the equation has only one solution, as it can be rewritten as $5(x - 5)^2 = 0$. And what does all this have to do with the number of fingers of Martians? Taking the mysterious question as a hint, we remember that we use the number of our fingers on both hands as the base of our number system. So, if our Martian had say b fingers, they would presumably use the base b number system, and read the equation as

$$5x^2 - 5bx + b^2 + 2b + 5 = 0$$

We also know that $x = 5$ is still a solution; this means that

$$b^2 - 23b + 130 = 0$$

which has the two solutions $b = 10$ and $b = 13$. We have already seen that for $b = 10$ there is only one solution, so we expect $b = 13$ to be the correct answer. In this case, the Martian version of the equation becomes, written in decimal numbers,

$$5x^2 - 65x + 200 = 0$$

which has indeed the two solutions $x = 5$ and $x = 8$.

Exercise 1.2.7 Find a second, different quadratic equation that would have stumped the Martian in exactly the same way.

Solution: The equation $3x^2 - 48x + 192 = 0$ has the only solution $x = 8$ when read decimally. But in base 13 it means $3x^2 - 60x + 288 = 0$, which has the second solution $x = 12$.

Exercise 1.2.8 What equation would you use for a Venusian, who is known to have 18 fingers?

Solution: The equation $7x^2 - 70x + 175 = 0$ has the only solution $x = 5$ when read decimally. But in base 18 it means $7x^2 - 126x + 455 = 0$, which has the second solution $x = 13$.

All algorithms from elementary school for addition, multiplication, and division of decimal numbers work with any number system.

Exercise 1.2.9 Find the periodic expansions of the fractions $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, $\frac{1}{5}$ and $\frac{1}{6}$ in base 7.

Solution:

$$\begin{aligned}\frac{1}{2} &= 0.\bar{4} \\ \frac{1}{3} &= 0.\bar{2} \\ \frac{1}{4} &= 0.\overline{26} \\ \frac{1}{5} &= 0.\overline{2365} \\ \frac{1}{6} &= 0.\bar{1}\end{aligned}$$

We are given a balance scale and can choose a set of four weights in multiples of a weight unit (let's say an ounce). What is the maximal number of weights one can determine with such a set?

Example 1.2.1 Suppose our weights are 1, 2, 4, and 8. This clearly allows to weigh and weight between 0 and 15 ounces, because any of these numbers is a sum of the first four powers of two. But because weights can be put on both pans of the scale, there are better solutions. Using weights of 1, 3, 9, and 27 ounces, we can weigh everything between 0 and 40 ounces.

More generally, using weights that are powers of 3, up to 3^n , we can weigh everything between 0 and $1 + 3 + \dots + 3^n$ ounces. To see this, write the weight in question as a number in base 3. For instance, to weight 35, we write $35 = 1022_3$. Now process this number from right to left: Whenever you encounter a 2, replace it with 3-1. In other words, replace the digit 2 by the 'digit' -1 and add 1 to the next digit to the left, using carriage if necessary. This way 1022_3 becomes successively $(1, 1, 0, -1)$, which means that $35 = 3^3 + 3^2 = 3^0$.

Exercise 1.2.10

1. Use the eight weights $1, 3, \dots, 3^7 = 2187$ to weigh 1627 ounces.
2. Is there a number that can be written in two different ways as a sum or difference of different powers of 3?
3. You have a set of six weights consisting of two weights each of 1, 5 and 25 ounces. What can you weigh with these weights?
4. You have a set of weights of 1, 3, and 10 ounces. The first weight you can't measure with these weights is 5 ounces. What additional weight would you need to add so that the first weight you can't measure is as large as possible?

1.3 THE GAME NIM

In many games, there are dedicated game pieces for each player. For instance, in chess White is not allowed to move with Black's pieces. We will now discuss games where this is not the case:

impartial

Definition 1.3.1 *An impartial game is a combinatorial game where the set of options depends only on the position but not on the player.*

This eliminates most games we commonly play, but leaves us with a class of games that can be well analyzed and yet offers challenging game play. Here is now the most fundamental of all impartial games, called NIM:

Two players have a few piles of tokens (like stones, matches, playing cards, coins) in front of them. They take turns taking away tokens: Each turn, the player picks a pile and removes any number (at least one) of tokens from that pile. She can take just one token, but she can also take the entire pile. Then it is the next player's turn. The player who takes the last token wins the game.

The game with just one pile is not very interesting: The first player wins by taking the entire pile.

Exercise 1.3.2 *Find a strategy for the game with just two piles.*

Solution: All positions where the two piles have different size are winning positions for the first player: She just needs to move so that both piles have the same size. All other position are losing positions.

Activity 1.3.3 *Play a game with 3 piles. Don't use too many tokens in a single pile.*

There is a simple way to characterize a winning position using binary numbers.

Suppose we have n piles of sizes $s_1, s_2, \dots, s_n$. Write the sizes s_n in binary form. For instance, if we have 3 piles of sizes 2, 5, 7, we would write

$$2_{10} = 010_2$$

$$5_{10} = 101_2$$

$$7_{10} = 111_2$$

Then we have:

Theorem 1.3.4 *A position is a losing position in NIM if every power of two occurs an even number of times in all pile sizes.*

Thus, our example is indeed a losing position, as the powers 2^0 , 2^1 , and 2^2 occur exactly twice, and all others not at all. If one writes down the pile sizes in binary form in a table as we did up above, one just needs to check whether every column in the binary representations of the numbers has an even number of 1's.

We will prove the theorem using two lemmas. Let's call a position *even* if every power of two occurs an even number of times in all pile sizes. In other words, if the pile sizes are given as binary numbers, we have an even number of 1's at every binary place.

Lemma 1.3.5 *Suppose a position is even. Then after any legal NIM move, the position cannot be even anymore.*

Proof. A legal NIM move changes exactly one of the numbers and therefore changes at least one of the binary digits. The corresponding power(s) of two will then occur an odd number of times. $\square$

Lemma 1.3.6 *Suppose a position is not even. Then there is a legal move that makes the position even.*

Proof. As the position is not even, there is a largest power of two that occurs an odd number of times. There is some pile in whose binary representation that power of two occurs. We change that pile size by choosing its binary digits so that all powers of two occur an even number of times in every column. Note that the leftmost digit we need to change is changed from 1 to 0, thus decreasing the pile size. $\square$

Both Lemmas together prove the theorem. Observe that the second Lemma gives us an effective strategy to find a winning move. Here is an example: Suppose the pile sizes are 2, 5, 6. To find a winning move, we write the numbers again in binary form:

$$2_{10} = 010_2$$

$$5_{10} = 101_2$$

$$6_{10} = 110_2$$

While 2^2 and 2^1 occur twice, 2^0 only occurs once, namely in the pile with 5 tokens. We need to change 101_2 into 100_2 by removing exactly one token from the second pile. This is the only winning move!

Exercise 1.3.7 *Find all winning moves for the following NIM positions:*

1. (3, 4, 5)
2. (3, 5, 7)
3. (1, 3, 5, 7)
4. (2, 4, 6, 8)
5. (2, 3, 5, 7, 11)
6. (1, 3, 5, 7, 9, 11)
7. (7, 8, 9, 10)
8. (8, 9, 10, 11)

Solution:

1. The second column is the only odd column, so we need to change the number 3 to 1 by taking two tokens from the first pile.
2. The last column is the only odd one. To make it even, we take a single token from any pile.
3. Here all columns are even. This is a losing position. If your opponent knows the strategy, she will beat you, no matter what you do.
4. $(2, 4, 6, 8) \mapsto (2, 4, 6, 0)$

5. $(1, 3, 5, 7, 9, 11) \mapsto (1, 1, 5, 7, 9, 11), (1, 3, 5, 5, 9, 11), (1, 3, 5, 7, 9, 9)$
6. $(7, 8, 9, 10) \mapsto (7, 4, 9, 10), (7, 8, 5, 10), (7, 8, 9, 6)$
7. $(8, 9, 10, 11)$: Losing position

Exercise 1.3.8 In a NIM position, one pile has at least size 10, while all others are of size at most 6. Show that this is a winning position.

Solution: The position can't be even, as there is only one pile with a 1 at binary position 2^3 or further left.

Exercise 1.3.9 Let's call a NIM position compelling if there is a winning move in every pile. If it's your turn to play in a compelling position, you can ask your opponent to point to the pile in which you move.

1. Show that $(4, 5, 6)$ is compelling.
2. Show that $(1, 5, 7, 9, 11)$ is compelling.
3. Find a number a so that $(2, 6, a, 10, 11)$ is compelling.
4. Is there a compelling position with four (non-empty) piles?
5. Find all compelling NIM positions with three piles up to size 7 where all piles have a different size.

Solution:

1. We have $G(3, 5, 6) = G(4, 2, 6) = G(4, 5, 1) = 0$.
2. We have $G(0, 5, 7, 9, 11) = G(1, 4, 7, 9, 11) = G(1, 5, 6, 9, 11) = G(1, 5, 7, 8, 11) = G(1, 5, 7, 9, 10) = 0$.
3. $a = 7$, because $G(0, 6, 7, 10, 11) = G(2, 4, 7, 10, 11) = G(2, 6, 5, 10, 11) = G(2, 6, 7, 8, 11) = G(2, 6, 7, 10, 9) = 0$.
4. No. Suppose we had such a position. Write the sizes in binary numbers. As the position is a winning position, there must be at least one power of 2 that occurs an odd number of times. Consider the largest such power. As there is an even number of piles, there is one pile where this power of 2 does not occur. By assumption, there must be a winning move that plays on this pile by reducing it. This move however cannot change the number of occurrences of the power of 2 we are contemplating.
5. $(4, 5, 6), (3, 5, 7), (4, 5, 7), (2, 6, 7), (3, 6, 7), (4, 6, 7), (5, 6, 7)$.

By introducing the NIM-sum of integers, the characterization of all winning positions becomes more transparent:

NIM-sum

Definition 1.3.10 Let m and n be integers. The NIM-sum $m \oplus n$ is the bitwise exclusive or addition of the binary representations of m and n . The exclusive or addition of the bits 0 and 1 is defined by

$\oplus$	0	1
0	0	1
1	1	0

Table 1.2 NIM-sum

Example 1.3.1 To compute $13 \oplus 26$, we write $13 = 01101_2$ and $26 = 11010_2$. The bitwise exclusive or addition yields 10111_2 , which represents the decimal number 23. Thus $13 \oplus 26 = 23$

$\oplus$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1	0	3	2	5	4	7	6	9	8	11	10	13	12	15	14	17	16	19	18	21
2	2	3	0	1	6	7	4	5	10	11	8	9	14	15	12	13	18	19	16	17	22
3	3	2	1	0	7	6	5	4	11	10	9	8	15	14	13	12	19	18	17	16	23
4	4	5	6	7	0	1	2	3	12	13	14	15	8	9	10	11	20	21	22	23	16
5	5	4	7	6	1	0	3	2	13	12	15	14	9	8	11	10	21	20	23	22	17
6	6	7	4	5	2	3	0	1	14	15	12	13	10	11	8	9	22	23	20	21	18
7	7	6	5	4	3	2	1	0	15	14	13	12	11	10	9	8	23	22	21	20	19
8	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	24	25	26	27	28
9	9	8	11	10	13	12	15	14	1	0	3	2	5	4	7	6	25	24	27	26	29
10	10	11	8	9	14	15	12	13	2	3	0	1	6	7	4	5	26	27	24	25	30
110	11	10	9	8	15	14	13	12	3	2	1	0	7	6	5	4	27	26	25	24	31
12	12	13	14	15	8	9	10	11	4	5	6	7	0	1	2	3	28	29	30	31	24
13	13	12	15	14	9	8	11	10	5	4	7	6	1	0	3	2	29	28	31	30	25
14	14	15	12	13	10	11	8	9	6	7	4	5	2	3	0	1	30	31	28	29	26
15	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0	31	30	29	28	27
16	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	0	1	2	3	4
17	17	16	19	18	21	20	23	22	25	24	27	26	29	28	31	30	1	0	3	2	5
18	18	19	16	17	22	23	20	21	26	27	24	25	30	31	28	29	2	3	0	1	6
19	19	18	17	16	23	22	21	20	27	26	25	24	31	30	29	28	3	2	1	0	7
20	20	21	22	23	16	17	18	19	28	29	30	31	24	25	26	27	4	5	6	7	0

Table 1.3 Nim addition values

Exercise 1.3.11 Show that $a \oplus b \leq a$ if and only if the leftmost 1 in the binary expansion of b is met by a 1 at the same position of the binary expansion of a .

Solution: If $b = 0$, the condition is empty (and thus trivially satisfied), and $a \oplus 0 = a$. Now let $b > 0$. If the leftmost 1 in the binary expansion of b is met by a 1 at the same position of the binary expansion of a , the corresponding digit in $a \oplus b$ becomes a 0, while all digits further left remain unchanged, thus decreasing the value. If the leftmost 1 in the binary expansion of b is met by a 0, the value of $a \oplus b$ is increased.

Remark 1.3.2 The integers $\mathbb{Z}$ together with NIM-addition form a group. Other groups you have encountered are the group $(\mathbb{Z}, +)$ of integers with the usual addition, or the group $(\mathbb{R}^+, \cdot)$ of positive real numbers with multiplication.

Proposition 1.3.12 A NIM position with piles of size $s_1, \dots, s_n$ is a losing position if and only if $s = s_1 \oplus s_2 \oplus \dots \oplus s_n = 0$. Then, the winning move consists of NIM-adding s to one of the s_i such that $s \oplus s_i < s$.

Let's try to visualize the set of losing positions in NIM. For the game with two piles of size x and y , we mark the coordinate square (x, y) in the (x, y) coordinate plane as green if the position is a losing position. As we have noted, the losing positions are exactly those where $y = x$, so the losing set is just the diagonal, as shown in **Figure 1.3**.

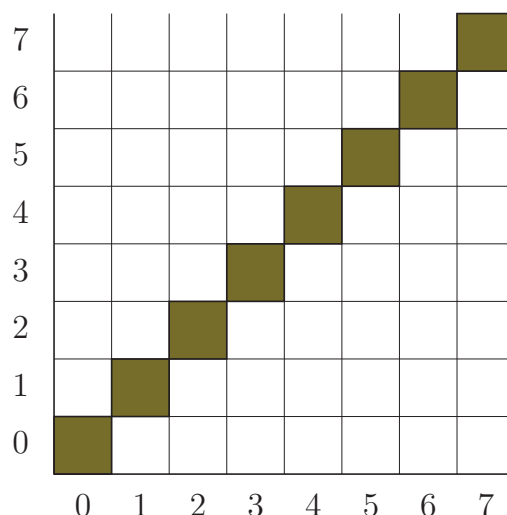


Figure 1.3 Losing positions
in NIM with two piles

The situation becomes dramatically more complicated and interesting if we switch to three piles of size x , y , and z . Marking the losing positions in the first octant in space, using (x, y, z) as coordinates, we obtain the images shown in **Figure 1.4**. There you can see the losing sets for pile sizes up to 1, 3, 7, and 15, respectively.

To understand what's going on, let's start with the simplest case where all piles are either empty or contain just one token. The losing positions are clearly those with an even number of tokens, as shown. Note that the cubes occur at the vertices of a tetrahedron. In the next case with pile sizes up to three, the shape of the simplest case repeats itself four times at the cube positions of the simplest case.

This has a simple explanation: Each cube represents all possible NIM positions with pile sizes being binary numbers with 2, 3, 4, or 5 digits. Let's partition the cubes in each image into eight subcubes of half-edge length each. In binary numbers, the coordinates of the points in the subcubes just differ in their leading places.

For instance, in the left-front-bottom subcube, all three coordinates have a leading 0, while in the right-back-top subcube, all coordinates have a leading 1.

In particular, we see that the left-front-bottom subcube of each picture must just be a copy of the previous picture, because it represents the same points, having an even number of 1's in each coordinate.

Now, the four empty subcubes in each picture with no losing positions are those where the three leading figures in the binary coordinates contain a single 1 and two 0s, making the leading column odd. The remaining three subcubes are exact copies of the left-front-bottom subcube, because changing two of the leading 0s to a 1 keeps the first column even.

This pattern repeats itself from image to image, creating a famous fractal called the *Sierpinski tetrahedron*.

Remark 1.3.3 Sometimes (for instance in ALAIN RESNAIS' famous film *Last Year in Marienbad*), NIM is played in the so-called *misère version*, where the player who takes the last token *loses*. There is a simple modification of the winning strategy for regular NIM, which works as follows: When in regular NIM the strategy asks you to move to a position where all piles are of size one, this would be an even number. In *misère NIM*, you want to make sure that there is an *odd* number of piles left. For instance, in the position $(4, 1, 1, 1)$, the first player wins *misère NIM* by taking all four tokens from the first pile.

There are a few modifications of the rules of NIM that also have very simple strategies.

Example 1.3.4 In one variation, we play as usual with a single pile of tokens, but are only allowed to remove one or two tokens from it. We use normal play, where the last player who moves wins.

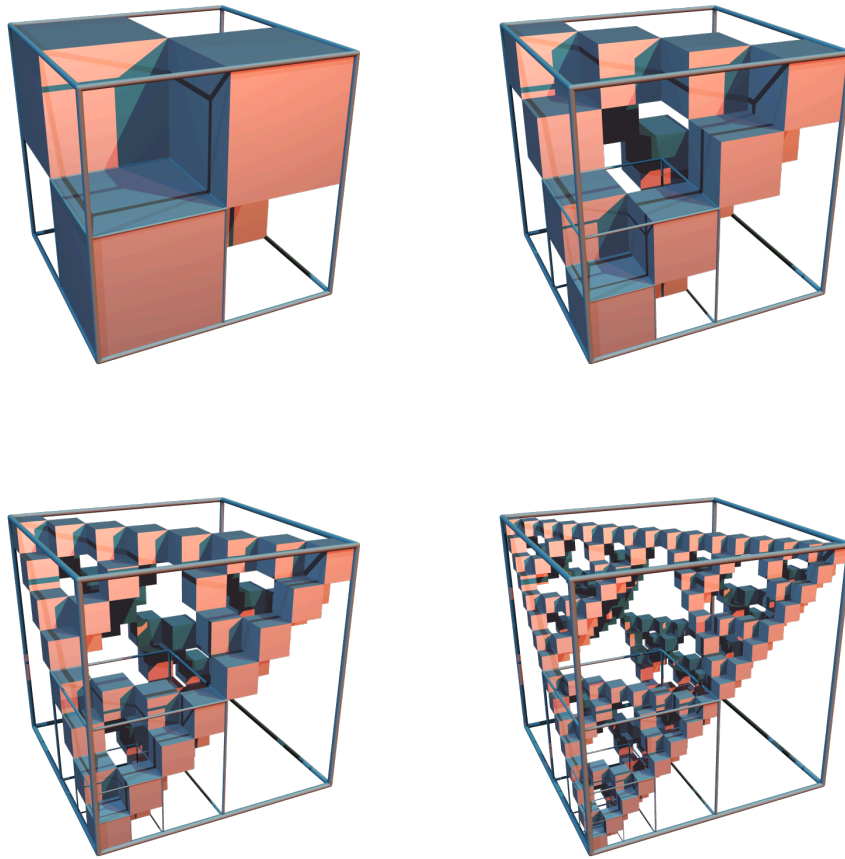


Figure 1.4 Losing positions in NIM with three piles

We claim that the first player can win if the size of the pile is not divisible by 3. She will move to reduce the pile size to a number divisible by 3. The second player must take one or two tokens, making the pile size indivisible by 3. From then on, the first player can continue with the strategy until the pile has reached size 0 and she wins. For example, the winning move for a pile of size 13 is to reduce it to size 12. No matter whether the second player reduces it to 11 or 10, the first player will answer by reducing it to 9, and so forth.

Exercise 1.3.13 Formulate a strategy for playing NIM with a single pile of tokens where each player can take one, two, or three tokens.

Solution: The strategy is to reduce the pile to a size divisible by 4.

We will discuss the more general case of playing these games with several piles in [Section 1.5](#) about Sprague-Grundy theory.

Exercise 1.3.14 You remember about a NIM position that one of the three piles had 3 tokens, and you could win by taking 2 tokens from either one of the two other piles. Show that the two other piles differed in size by 1.

Solution: Denote the sizes of the unknown piles by x and y . Then the assumptions are that $3 \oplus (x - 2) \oplus y = 0$ and $3 \oplus x \oplus (y - 2) = 0$. Making a case distinction mod 4, the first assumption implies

$$x = \begin{cases} 4k \\ 4k + 1 \\ 4k + 2 \\ 4k + 3 \end{cases} \Rightarrow y = \begin{cases} 4k - 3 \\ 4k - 5 \\ 4k + 3 \\ 4k + 2 \end{cases}$$

Together with the symmetric second assumption, we have $\{x, y\} = \{4k + 2, 4k + 3\}$ for some integer k

1.4 NIM IN DISGUISE

In this section, we discuss a few games that are disguised versions of NIM. Most of these games are taken from [2].

Example 1.4.1 (POKER NIM) is played like NIM with a number of piles of tokens, except that one (or both) players have their own (finite) reservoir of tokens. A move is either a normal NIM move, reducing the number of tokens in a single pile, or consists of *increasing* the number of tokens in a single pile, using one's reservoir.

A game position is a winning position if and only if that's the case for regular NIM. The winning strategy is also the same as for NIM, with the addendum that any move from the reservoir is countered by removing the coins that were just added from the same pile. This way, the position remains the same except that the reservoir got smaller.

Exercise 1.4.1 You have a reservoir of five tokens and have decided to use some of your tokens to increase the size of one pile in POKER NIM. For the following pile sizes, which pile(s) should you not increase?

1. (1, 4, 7)
2. (2, 3, 4)
3. (2, 4, 7)

Solution:

1. You do not want to increase pile one by 2 tokens or pile two by 2 tokens.
2. You do not want to increase pile one by 5 tokens or pile two by 3 tokens.
3. You do not want to increase pile one by 1 token or pile two by 1 token.

Example 1.4.2 (SLIDING) is another game that uses the winning strategy we discovered for NIM. It is being played on a horizontal strip of n squares, with coins placed on some squares. A move consists of moving any one coin to the right across empty squares onto an empty square. In other words, the coin may occupy any one of its empty neighbors to the right, up to the next occupied spot.

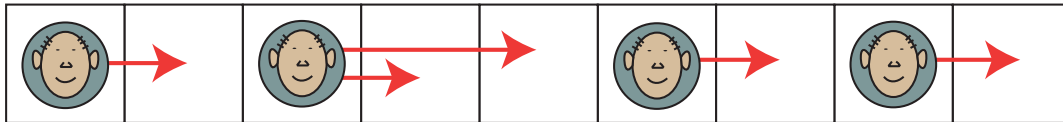


Figure 1.5 Possible moves for SLIDING

Here, the NIM piles are disguised as gaps between coins — but only every other gap counts. We denote the lengths of the gaps between coins by $g_1, g_2, \dots, g_n$, with the leftmost gap being g_1 , and the rightmost gap at the end of the strip. In our example, this sequence would be 1, 2, 1, 1. A SLIDING position is equivalent to NIM piles of the size of the *odd* gaps $g_1, g_3, \dots$.

To see this, number the coins from left to right as 1, 2, ... Observe that the gap g_i occurs between coin i and $i + 1$. Moving an odd coin decreases the odd gap to the right of the coin, while moving an even coin increases the odd gap to the left of the coin. The former is a standard move on the NIM pile, while the latter is countered by decreasing the increased gap to its former size by moving the odd coin to the left of the gap.

Exercise 1.4.2 Number the squares of the strip from left to right, starting with 1. We denote a SLIDING position with coins on the squares $n_1 < n_2 \dots < n_k$ for a board of total width n just by $(n : n_1, n_2 \dots, n_k)$. Find all winning moves for the following positions:

1. $(14 : 1, 3, 6, 9, 10)$
2. $(22 : 1, 4, 8, 12, 14, 19)$
3. $(30 : 1, 3, 8, 13, 17, 25)$

Solution:

1. This game corresponds to the NIM position $(1, 2, 4)$ with only winning move to $(1, 2, 3)$. This move corresponds to moving the coin on square 10 to square 11. None of the gap-increasing moves are winning moves.
2. This game corresponds to the NIM position $(2, 3, 4)$ with only winning move to $(2, 3, 1)$. This move corresponds to moving the coin on square 14 to square 17. None of the gap-increasing moves are winning moves.
3. This game corresponds to the NIM position $(1, 4, 7)$ with only winning move to $(1, 4, 5)$. This move corresponds to moving the coin on square 17 to square 19. Among the distance-increasing moves, the winning moves are to move 3 to 5 or 13 to 15.

Example 1.4.3

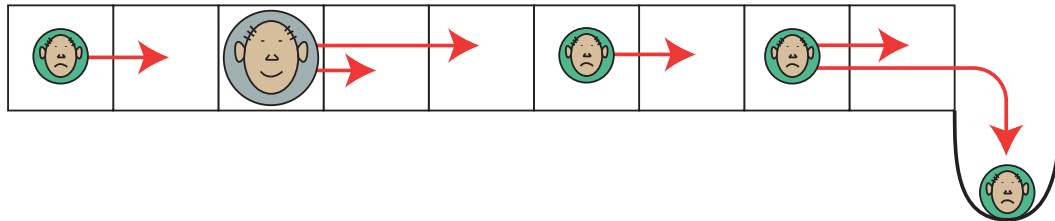


Figure 1.6 Possible Moves

The SILVER DOLLAR GAME is played like SLIDING on a rectangular strip. One of the coins is a silver dollar, and the others are worthless pennies. The rightmost field is designated as the money bag. Again, both players take turns sliding a single coin towards the right across any number of empty fields. When a coin reaches the money bag, it is being pocketed by the player who moved it. The player who captures the silver dollar wins. You can see all possible moves indicated by arrows in **Figure 1.6**.

To discuss this game, we look at two simple cases first:

If there is no penny to the right of the silver dollar, the obvious winning move is to slide the silver dollar into the money bag, as in **Figure 1.7**.

In the case where there is just one penny to the right of the money bag, the player who moves this penny into the money bag loses, because the next player will pocket the silver dollar. Thus, this case of the game is like playing SLIDING on a board without the money bag: If we can win the corresponding SLIDING game, the game will end with our opponent being forced to move the right penny into the money bag, and we can grab the silver dollar.

Example 1.4.4 In **Figure 1.8**, the sizes of the odd gaps are $g_1 = 0$, $g_2 = 2$, and $g_5 = 3$. The NIM-sum of these numbers is 1, hence a winning requires changing a gap size by NIM-adding 1 to it. This would result in new gap sizes $\tilde{g}_1 = 1$,

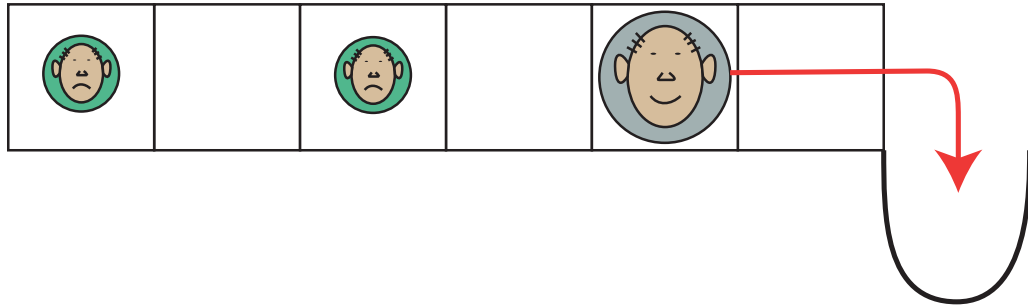


Figure 1.7 Easy Money

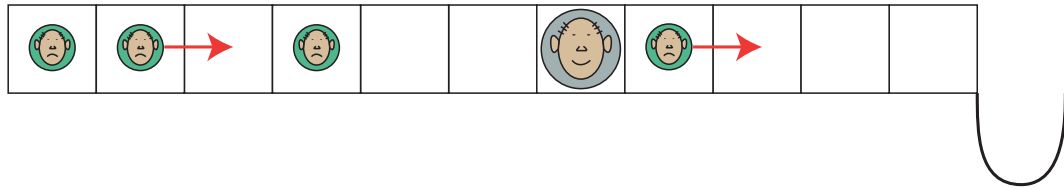


Figure 1.8 One penny matters

$\tilde{g}_2 = 3$, and $\tilde{g}_3 = 2$. The first and last can be realized by moving the penny as indicated. If there were another empty field to the right of the silver dollar, we could also increase the second gap to size 3 by moving the silver dollar.

If there are at least two pennies to the right of the silver dollar, moving a penny into the money bag is not necessarily catastrophic. We claim:

Lemma 1.4.3 *A position with at least two pennies to the right of the silver dollar is a winning position if and only if it is a winning position for the corresponding SLIDING position, where now the money bag is included as a field.*

Hence: Whenever we have a position with at least one penny to the right of the silver dollar, there is a corresponding position of SLIDING. In case there is only one penny to the right of the silver dollar, we exclude the money bag from the bad count; otherwise, we include it.

Proof. Let's call a position of the SILVER DOLLAR GAME where the corresponding SLIDING position is a winning position a *W*-position, and any other position an *L*-position. Here, we include the money bag on the board if there are at least two pennies to the right of the silver dollar, as explained.

To prove the lemma, we need to show two statements:

1. In a *W*-position, there is a move that turns the game into an *L*-position.
2. In an *L*-position, every possible move leads to a *W*-position.

We have seen this already if there is only one penny to the right of the money bag, so from now on we assume there are at least two.

The main difficulty in the argument lies in the possibility that we are sometimes allowed to remove pennies and to shorten the board, which leads to some case distinctions.

Let's first assume that we have a *W*-position. Thus, the corresponding SLIDING position is a winning position, which we can always win by decreasing the gap size, as we are playing NIM with the sizes of the odd gaps. The

new position is an L -position for the SILVER DOLLAR GAME, even if we move a penny into the money bag. (Note that there might be a winning move for the SLIDING position that increases a gap, moves a coin in the money bag, and still has at least two coins to the right of the silver dollar. That position will not be an L position, as the new gap size after removing the penny is one larger than with a penny.)

In the second case, we assume that we have an L -position for the SILVER DOLLAR GAME. We have to show that every move turns this position into a W -position. If the move doesn't put a penny into the money bag, this assertion is true, as we are just playing SLIDING. Hence, suppose our move moves the rightmost penny into the money bag. There are two cases: In case the last gap is an odd gap, this move reduces it to size 0, and the resulting SLIDING position has non-zero NIM-sum. The same holds for the SLIDING position after removing the penny and possibly the money bag, as doing so doesn't affect the remaining odd gaps. In case the second last gap is odd, this move increases that gap size. If we remove the penny and the money bag, the new last gap size remains the same, or increases even further by 1, if we keep the money bag as a field. In either case, the resulting position has a non-zero NIM sum for the odd gaps, and is therefore a W -position, which we had to show.

Example 1.4.5

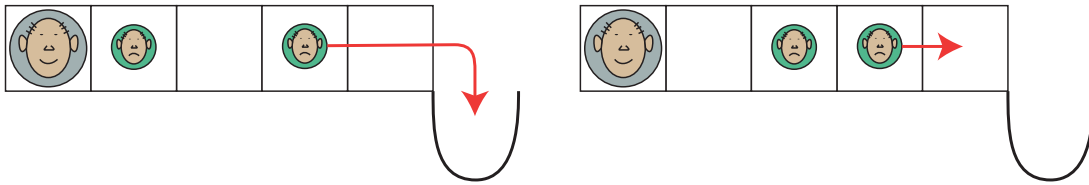


Figure 1.9 Two pennies to the right of the silver dollar

In **Figure 1.9**, there are two pennies to the right of the silver dollar. In the first case, the gap sizes of the odd gaps are 0 and 2. To win, we need to make them both equal, and the only possibility is to decrease the second gap to size 0. This requires moving the right penny into the money bag, thus eliminating the coin, the money bag, and the second odd gap.

Example 1.4.6

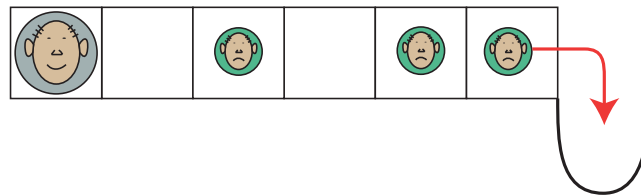


Figure 1.10 A dangerous move

In **Figure 1.10**, there are three coins to the right of the silver dollar, so we have to include the money bag in the gap count. The gap sizes are $g_1 = g_2 = 1$, $g_3 = 0$, and $g_4 = 1$. This is a winning position for the first player, and the correct winning move is to decrease the first gap to 0 by moving the silver dollar one position to the right. It seems that there is another winning move, indicated by the arrow, that would increase the third gap to size 1. That, however, is not possible, as after sliding the right coin in the money bag, it disappears, leaving it with a gap of size 2 instead of size 1.

If, on the other hand, the first gap in this example has size 2, then moving the right coin into the money bag would indeed be one of the two winning moves!

Exercise 1.4.4 Find a position for the SILVER DOLLAR GAME with three pennies to the right of the silver dollar such that moving the right penny into the money bag is a winning move.

Solution: The easiest way to find a solution is to find a position with *two* pennies to the right of the money bag that is a *losing* position, and then put another penny to the right of the others. An example is shown in Figure 1.11.

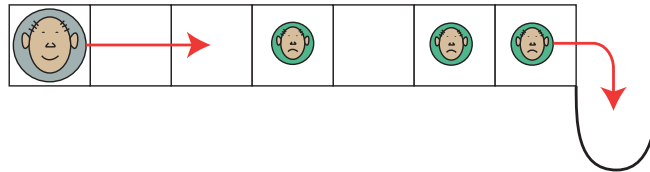


Figure 1.11 Moving into the money bag wins

Exercise 1.4.5 Suppose it's your opponent's turn in the position in Figure 1.6, and she moves the rightmost penny into the money bag. How do you respond?

Solution: After her move, the position corresponds to the NIM position $(1, 3)$, with the pile sizes corresponding to the gap between the leftmost penny to the silver dollar, and the 3 corresponding to the gap between the penny to the right of the silver dollar and the right end of the strip. The winning move is to reduce the second pile to height 1. This means that we need to move the penny to the right of the silver dollar two squares to the right.

Exercise 1.4.6 In the SILVER DOLLAR GAME, label all fields from left to right with 1, 23, ..., 20, so that the money bag is to the right of the field with number 20. Find all winning moves (if any) in the following positions

1. Pennies at 1, 11, 17, Silver Dollar at 3;
2. Pennies at 1, 6, 14, 16, Silver Dollar at 9;
3. Pennies at 3, 5, 7, 15, 19, Silver Dollar at 1.

Solution:

1. Move to 1, 7, 11, 17 or to 1, 3, 15, 17.
2. Move to 4, 6, 9, 14, 16, to 1, 6, 12, 14, 16, or to 1, 6, 9, 14, 21.
3. Move to 1, 4, 5, 7, 15, 19 or 1, 3, 5, 8, 15, 19, or 1, 3, 5, 7, 18, 19.

Exercise 1.4.7 In the SILVER DOLLAR GAME, label all fields from left to right with 1, 23, ..., 20, so that the money bag is to the right of the field with number 20. The pennies are at 1, 3, 7, 12, 19. Where can you place the Silver Dollar so that the position becomes a losing position?

Solution: Place the Silver Dollar at position 13.

Exercise 1.4.8 In the SILVER DOLLAR GAME, label all fields from left to right with 1, 2, 3, ... The pennies are at positions 3, 6, 9, and 13, and the Silver Dollar is at position 1. How long must the board be (without the money bag) so that this is a losing position?

Solution: The board must have a length of 15.

Example 1.4.7 NORTHCOTT'S GAME is played on an $m \times n$ chess board, with one purple and one orange checker in each row. The two players, Purple and Orange, can move one checker of their color by one field to the left or right (but not vertically) on an empty field.

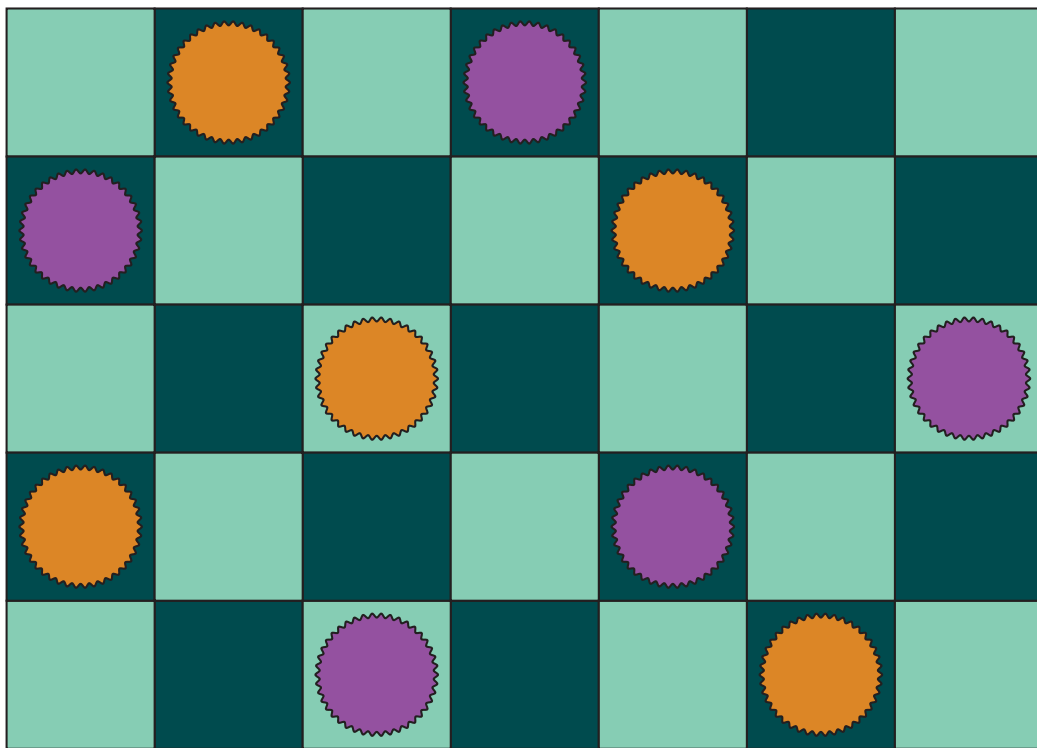


Figure 1.12 NORTHCOTT'S GAME

Here, the sizes of the NIM piles are the gaps between the checkers in each row. If a player increases a gap, this is countered by decreasing the gap to its previous size.

Exercise 1.4.9 Determine all winning moves for the position shown in **Figure 1.12**, if any.

Solution: The NORTHCOTT position corresponds to the NIM position $(1, 3, 3, 3, 2)$, which is a losing position.

Example 1.4.8 RIMS is played on a sheet of paper with some dots and loops separating the dots. A move consists of drawing a closed loop through a positive number of dots. The loop may neither cross itself nor any other loop.

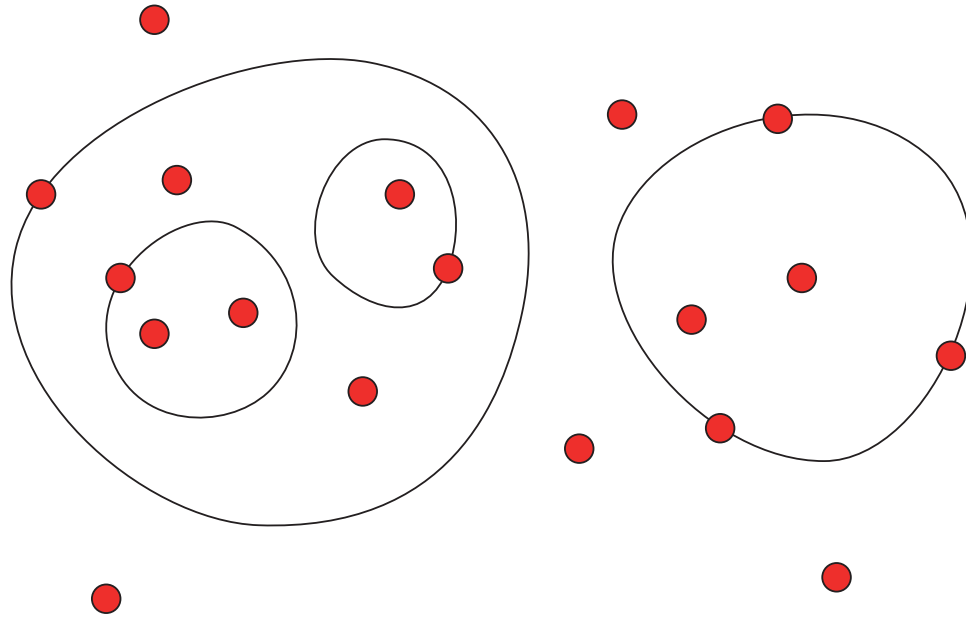


Figure 1.13 RIMS

The loops that are on the paper separate the plane into r regions, each of which may contain a certain number n_i , $i = 1, \dots, r$ of dots. A move will create a new loop within one of these components, thereby reducing the number of dots by any number between 1 and n_i , and splitting the remaining number m_i of dots into a sum $m_i = n'_i + n''_i$ with non-negative n'_i, n''_i . Instead of using dots and arcs, we can just as well play a modification of NIM that allows us to split a pile into two after it has been reduced. As any such move can be realized by drawing a suitable arc, these two versions of the game are indeed equivalent.

Like normal NIM, this modified version can be won exactly the same way. To see this, it suffices to show that if we are in a losing position, any legal move will change this to a winning position (for the other player).

Recall that a losing position is one where the NIM sum of the piles is 0. A legal move consists of replacing a single pile by possibly more than one pile whose total size is strictly less than the size of the removed pile. As the NIM sum can be perceived as the usual sum without carrying, the NIM sum of the replacing piles is less than or equal to their actual total size, and thus not equal to the size of the old pile.

Thus, the new position has a non-zero NIM sum and is a winning position for the other player.

Exercise 1.4.10 Find all winning moves for the position in Figure 1.13, if any.

Solution: The RIMS configuration corresponds to the NIM configuration $(5, 2, 2, 2, 1)$ which has as its only winning move to reduce the size of the pile 5 to size 3. Thus, we need to draw a loop through two of the dots in the outside component. It does not matter whether and how we separate the remaining 3 dots.

Exercise 1.4.11 *The game BABEL is played on a rectangular strip like SLIDING. Each field can have a tower of several coins on it. At each turn, the player moves one or more coins from one tower to the field to its right, adding to the tower if one is already present there. The player who completes the Tower of Babel at the right end of the game board wins. Explain how this game is another version of NIM in disguise. Hint: Label the fields from right to left with the numbers 0, 1, 2, etc, and pay attention to the even and odd fields.*

Solution: We ignore all even fields and consider the odd ones as NIM piles. Moving coins from an odd field to the right corresponds to removing the coins from that NIM pile. If our opponent moves from an even pile, this corresponds to adding coins to the NIM pile to the right. We counter that by moving the same coins she just moved one position further to the right.

1.5 SPRAGUE-GRUNDY THEORY

In this section, we study a special class of games called *impartial* games that generalize NIM.

We will first assign a number to each position in an impartial game, its *Grundy number* $\mathcal{G}$. For this, we will need the mex-function.

mex

Definition 1.5.1 *The mex-function assigns to any finite set of non-negative integers the smallest non-negative integer that is not in the set.*

Example 1.5.1 For instance, $\text{mex}(\{\}) = 0$ and $\text{mex}(\{1, 2, 3\}) = 0$, but $\text{mex}(\{0, 2, 4\}) = 1$. The name mex stands for minimal exclusive number.

$\mathcal{G}$

Definition 1.5.2 *Consider a position in an impartial game. The player has a finite set of possible moves, leading to a set of new positions. We define the Grundy number $\mathcal{G}$ of the position as 0 if the set of possible moves is empty (i.e., the player has lost). Otherwise, the Grundy number $\mathcal{G}$ of the position is the mex of the set of Grundy numbers of all positions that can be reached by a single legal move.*

This is a so-called *recursive* or *inductive* definition. We have to convince ourselves that this definition makes sense. Because we insisted that impartial games are finite, every position has only a finite set of possible subsequent positions. As the Grundy numbers for the end positions are defined to be 0, we can compute the value of all the second last positions — they are necessarily $1 = \text{mex}(0)$. Then we compute the Grundy number of the third last positions using the mex-function. It can be 0 or 2. As the definition of the Grundy number of a position only involves later values, we can work our way backwards to the position we are actually interested in.

Let's do a simple example:

Proposition 1.5.3 *The Grundy number of a NIM-pile with n tokens is n .*

Proof. Denote the NIM-pile of size n by N_n . We prove $\mathcal{G}(N_n) = n$ by induction. If $n = 0$, there is no move left, and then $\mathcal{G}(N_0) = 0$. Now suppose we know that $\mathcal{G}(N_k) = k$ for all $k \leq n$. We need to show that $\mathcal{G}(N_{n+1}) = n + 1$. The possible moves reduce the NIM-pile of size $n + 1$ to all possible sizes between 0 and n , so that

$$\mathcal{G}(N_{n+1}) = \text{mex}(\{0, 1, \dots, n\}) = n + 1$$

as desired. □

The Grundy number of a position tells us whether it is a winning or losing position, because

Proposition 1.5.4 *A game G is in a winning position if and only if $\mathcal{G}(G) > 0$.*

Proof. By the definition of the mex-function,

1. for a position with $\mathcal{G} > 0$, there is always an option to a position with $\mathcal{G} = 0$;
2. for a position with $\mathcal{G} = 0$, all options have $\mathcal{G} > 0$.

This implies that a winning strategy is to always move to a position with $\mathcal{G} = 0$. □

But the Grundy number of a game contains more information than just whether it's a losing or winning position. This becomes relevant when we consider the *sum* of games:

Definition 1.5.5 The sum $G_1 + G_2 + \cdots + G_n$ of games $G_1, \dots, G_n$ is defined by allowing each player to first select one of the games G_i , and then take a move in that game.

For instance, the general NIM game is just the sum of the single pile NIM games consisting of the individual piles.

Exercise 1.5.6 If G is any impartial game, show that $\mathcal{G}(G + G) = 0$.

Solution: The second player has a winning strategy by using symmetric play. Thus, the claim follows from **Proposition 1.5.4**.

Next, we will compute the Grundy number of the general NIM game.

Proposition 1.5.7 Let G_i be a NIM-pile of size N_i . Then

$$\mathcal{G}(G_1 + G_2 + \cdots + G_n) = N_1 \oplus N_2 \oplus \cdots \oplus N_n$$

Proof. We prove this by induction over the sizes of all piles. If all but one pile are empty, the claim follows from **Proposition 1.5.3**. Now assume that G_i is a pile of size N_i , and the claim has been proven for all NIM games with smaller piles. Let $N = N_1 \oplus N_2 \oplus \cdots \oplus N_n$. We have to show

1. For every integer $k = 0, 1, \dots, N - 1$, there is an option of $G_1 + G_2 + \cdots + G_n$ with Grundy number k ;
2. There is no option of $G_1 + G_2 + \cdots + G_n$ with Grundy number N .

To see this, write the N_i in binary form. We would like to change the NIM-sum $N = N_1 \oplus N_2 \oplus \cdots \oplus N_n$ to k by decreasing one of the N_i . To do so, we look at the binary expressions for $N \oplus k$ and note the bits that are equal to 1. If we change the corresponding bits in *any* of the N_i , the NIM-sum becomes equal to k . We need to show that there is an N_i that is decreased this way. Consider the leftmost (highest) position of the bits we need to change.

We need to find an N_i that has a 1 at the same position. If that doesn't exist, then N also has a 0 at this very position, and NIM-adding $N \oplus k$ to N would result in a number larger than N , contradicting that $N \oplus k \oplus N = k < N$.

This shows that we can decrease one N_i so that the NIM-sum becomes k . By our induction hypothesis, this NIM-sum equals the Grundy number of the new NIM position, which proves the first claim. For the second, suppose we can reduce one pile so that the Grundy number remains the same, say N . Again by induction, it suffices to show that the NIM-sum changes whenever we make a move. But this is obvious, as a move changes at least one bit in a single pile. $\square$

Example 1.5.2 We illustrate this proof with an example. Suppose that we have two NIM-piles of sizes $6 = 0110_2$ and $11 = 1011_2$. Then $6 \oplus 11 = 1101_2 = 13$, and we claim that this number is the Grundy number of the game. Clearly, any move will change the NIM-sum so that (by induction) there is no move to a new position with Grundy number 13. On the other hand, for all smaller numbers, we can find a move so that the resulting position has this number as its NIM-sum. For instance, if we want the NIM-sum to be $12 = 1100_2$, we need to NIM-add $12 \oplus 13 = 0001_2$ to either number *and* make sure that the value decreases. This works for 11, as $11 \oplus 1 = 10$.

The surprising discovery of ROLAND SPRAGUE and PATRICK GRUNDY was that the same formula holds not just for NIM piles but for any sum of impartial games:

Theorem 1.5.8 For impartial games $G_1, G_2, \dots, G_n$,

$$\mathcal{G}(G_1 + G_2 + \dots + G_n) = \mathcal{G}(G_1) \oplus \mathcal{G}(G_2) \oplus \dots \oplus \mathcal{G}(G_n)$$

Proof. Let $g = \mathcal{G}(G_1) \oplus \mathcal{G}(G_2) \oplus \dots \oplus \mathcal{G}(G_n)$. We have to show that

1. for every integer $k = 0, 1, \dots, g - 1$, there is an option of $G_1 + G_2 + \dots + G_n$ with Grundy number k ;
2. there is no option of $G_1 + G_2 + \dots + G_n$ with Grundy number g .

We argue again by induction, this time over the number of moves that are left until the game ends. If this number is 0, there are no options in any game, and both the Grundy numbers of all G_i and of their sum are 0. Now, assume that we have a more engaging state of the game, and the formula has been proven for all later positions.

To find an option for the first claim, pretend the G_i are NIM-piles of sizes $N_i = \mathcal{G}(G_i)$. By **Proposition 1.5.7** we can change the Grundy number of the NIM-sum $N_1 \oplus \dots \oplus N_n$ to g by decreasing one NIM-pile. Let's say that we need to decrease N_j to $N'_j < N_j$. As the corresponding impartial game G_j has Grundy number N_j , there is an option in G_j that decreases its Grundy number to N'_j . By induction hypothesis, the Grundy number of the new position equals k . To see that there is no option in the second case, assume there was a move that leaves the Grundy number unchanged, say by reducing the Grundy number of game G_j to a number $N'_j < N_j = \mathcal{G}(G_j)$.

Consider the corresponding game of NIM piles of sizes $N_i = \mathcal{G}(G_i)$. By induction hypothesis, reducing the pile N_j to size N'_j would leave the NIM-sum of the N_i unchanged, contradicting **Proposition 1.5.7**. $\square$

Thus, *any* impartial game is equivalent to a NIM pile of some size. As we know how to play NIM, the important part is to determine the size of the NIM pile.

As an application, let's play the variation of NIM from **Example 1.3.4** with several piles: We are given a collection of r piles of tokens of size $n_1, \dots, n_r$, and a move consists of removing one up to $k - 1$ tokens from a single pile. Let $N(k, n)$ denote a game with a pile of size n where it is allowed to take up to $k - 1$ tokens from that pile.

Lemma 1.5.9 The Grundy number of the single pile game $N(k, n)$ equals $n \bmod k$.

Proof. First, all pile sizes of the form $k, 2k, 3k$, etc have Grundy number 0, because the second player has the following winning strategy: If the first player takes s tokens, the second takes $k - s$, keeping the size of the pile divisible by k . Secondly, the piles of size $n = 1, 2, \dots, k - 1$ have Grundy number n , as the game is no different from the normal NIM for small piles. Thirdly, for $n > k$ we proceed by induction. For a pile of size n , we can assume that smaller piles of size $n' < n$ have Grundy number $n' \bmod k$. As we can take away up to $k - 1$ tokens, the resulting $k - 1$ piles will have as Grundy numbers all numbers from 0 to $k - 1$ except the value $n' \bmod k$. $\square$

Corollary 1.5.10 Then

$$\begin{aligned} \mathcal{G}(N(k_1, n_1) + \dots + N(k_r, n_r)) &= \mathcal{G}(N(k_1, n_1)) \oplus \dots \oplus \mathcal{G}(N(k_r, n_r)) \\ &= (n_1 \bmod k_1) \oplus \dots \oplus (n_r \bmod k_r) \end{aligned}$$

Proof. Clear. $\square$

Exercise 1.5.11 Find all winning moves in the following games, if any:

1. $N(2, 10) + N(3, 10) + N(4, 10)$
2. $N(5, 11) + N(5, 13) + N(5, 17)$

Solution:

1. $\mathcal{G}(N(2, 10) + N(3, 10) + N(4, 10)) = 3$, so this is a winning position. Winning moves either reduce the second pile to 2, 5, 8 tokens or the third pile to 1, 5, 9 tokens.
2. $\mathcal{G}(N(5, 11) + N(5, 13) + N(5, 17)) = 0$, so this is a losing position.

Exercise 1.5.12 This game is played with a single pile of size n . If n is even, you are allowed to take 1 or 3 tokens, and when n is odd, you can take 1 or 2 tokens. Determine the Grundy value for piles of size 0 through 16 until you discover a pattern.

Solution: The Grundy values are 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0.

Example 1.5.3 In LASKER'S NIM, we play with piles as usual. A move either consists of removing any positive number of tokens from a single pile, or of dividing one pile into two smaller piles, without taking away any tokens.

We compute the Grundy function for the single pile game as follows: Clearly, $\mathcal{G}(0) = 0$ and $\mathcal{G}(1) = 1$. The pile of size 2 has the options 0, 1, or (1, 1) which have Grundy numbers 0, 1 and $1 \oplus 1 = 0$, so that $\mathcal{G}(2) = 2$.

For general n , we have the recursion formula

$$\mathcal{G}(n) = \text{mex}(\mathcal{G}(1), \dots, \mathcal{G}(n-1), \\ \mathcal{G}(1) \oplus \mathcal{G}(n-1), \mathcal{G}(2) \oplus \mathcal{G}(n-2), \dots, \mathcal{G}(n-1) \oplus \mathcal{G}(1)).$$

Using it, we compute

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\mathcal{G}(n)$	0	1	2	4	3	5	6	8	7	9	10	12	11	13	14	16

Table 1.4 Grundy numbers for LASKER'S NIM

This suggests the general formula

$$\mathcal{G}(n) = \begin{cases} 4k+1 & \text{if } n = 4k+1 \\ 4k+2 & \text{if } n = 4k+2 \\ 4k+4 & \text{if } n = 4k+3 \\ 4k+3 & \text{if } n = 4k+4 \end{cases}$$

for all $k \geq 0$. We can prove this formula by induction, using the recursion formula. For instance, if $n = 4k+3$, then the single-pile options of n have Grundy numbers $0, \dots, 4k+2$. Hence, the Grundy number of n is at least $4k+3$. The 2-pile options $(1, 4k+2)$ have Grundy number $1 \oplus 4k+2 = 4k+3$, and all other 2-pile options have a smaller Grundy number. This shows that $\mathcal{G}(4k+3) = 4k+4$. The other cases are dealt with likewise.

Exercise 1.5.13 In GREEDY NIM, a player must take at least half of the tokens from a single pile. Show that the Grundy number of a game with a single pile of size n is the number of digits in the binary expansion of n .

Solution: Let g_n be the Grundy number of a GREEDY NIM game with a single pile of size n . Then $g_0 = 0$, $g_1 = 1$ and $g_2 = \text{mex}\{g_0, g_1\} = 2$. We prove the claim by induction over the number k of digits in the binary expansion of n . Suppose the claim is true for all n that have at most k binary digits. Let N be a number with $k + 1$ binary digits. Thus $2^k \leq N < 2^{k+1}$. As in a single move the pile is at least halved, the possible pile sizes after that move are all non-negative integers up to a number N' with $2^{k-1} \leq N' < 2^k$. These numbers have between 0 and k binary digits. By induction hypothesis, their Grundy numbers form the set $\{0, \dots, k\}$ so that $g_N = \text{mex}(0, \dots, k) = k + 1$ as claimed.

Exercise 1.5.14 In the game TOUCH-ME-NOT, both players take turns placing tokens on the empty squares of a strip of squares of length n so that two tokens are never on neighboring squares.

1. Compute the Grundy numbers for all positions with up to 7 squares. For example, you should find that $G(7) = 1$.
2. Find all four winning moves for the empty strip with 6 squares.

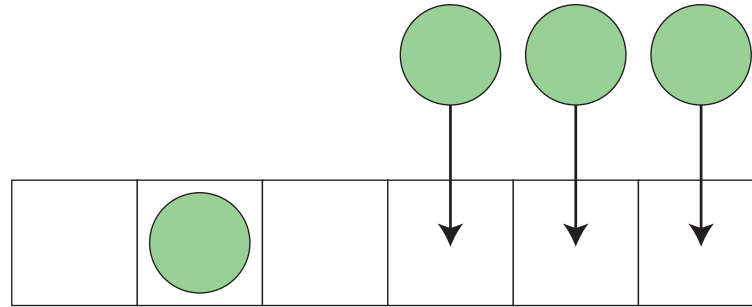


Figure 1.14 Legal moves in TOUCH-ME-NOT

Solution:

1. Denote by g_n the Grundy number of TOUCH-ME-NOT for an empty strip of n squares. Then we have the recursion

$$g_n = \text{mex}(g_{n-2}, g_0 \oplus g_{n-3}, g_1 \oplus g_{n-4}, \dots, g_{n-3} \oplus g_0)$$

which leads to

n	1	2	3	4	5	6	7	8	9	10	11	12
g_n	1	1	2	0	3	1	1	0	3	3	2	2

Table 1.5 Grundy numbers for TOUCH-ME-NOT

This sequence becomes eventually ($n > 51$) periodic with a period of length 34.

2. To win, the token needs to be placed on field 1, 3, 4 or 6.

1.6 MORE EXAMPLES OF IMPARTIAL GAMES

While we have seen that impartial games are equivalent to a single NIM pile, this only really helps if we can determine the size of the NIM pile, i.e., the Grundy number of the game. In this section, we are looking at some games where the Grundy numbers can be computed quite nicely to the extent that the games become actually playable.

Exploration 1.6.1 In the game MAGNETISM, both players take turns placing $\oplus$ and $\ominus$ tokens on the empty squares of a strip of squares of length n so that two tokens of the same polarity are never on neighboring squares. Both players can use tokens of either polarity. In the following problems, you will compute the essential Grundy numbers of this game. Any game position is equivalent to a game sum of the positions discussed below. For all four of the Grundy sequences you will compute below, expect very simple patterns.

1. Compute the Grundy numbers for all positions on strips with up to 6 empty squares.
2. Compute the Grundy numbers for all positions on strips with up to 6 squares with a $\oplus$ token on the leftmost square.
3. Compute the Grundy numbers for all positions on strips with up to 6 squares with a $\oplus$ token on the leftmost square and a $\oplus$ token on the rightmost square.
4. Compute the Grundy numbers for all positions on strips with up to 6 squares with a $\oplus$ token on the leftmost square and a $\ominus$ token on the rightmost square.
5. Use induction to prove the observations above.
6. Find all winning moves for the strip with 7 squares that has a $\oplus$ token on the leftmost square.
7. Discuss the same game on boards with different topology. For instance, you board can have loops and branches.

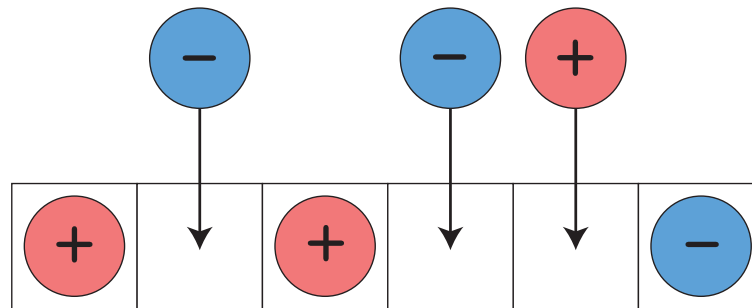


Figure 1.15 Legal moves in MAGNETISM

Solution: Denote by g_n the Grundy number of MAGNETISM for an empty strip of n squares, by g_n^+ the Grundy number for a strip of n squares with a $+$ token on the leftmost square, by g_n^{++} the Grundy number for a strip of n squares with a $+$ token on the leftmost square and a $+$ token on the rightmost square, and by g_n^{+-} the Grundy number for a strip of n squares with a $+$ token on the leftmost square and a $-$ token on the rightmost square.

Then we have

All sequences continue like this, as can be proven by induction. To win on the strip with 7 squares that has a $\oplus$ token on the leftmost square, either place a $+$ token on square 6 or a $-$ token on square 7.

TETRIAMOND is played by two players taking turns on a triangular grid, as shown, that lies above a floor line (bold), with a tetriamond (see Section 2.2) placed somewhere above ground.

n	1	2	3	4	5	6
g_n	1	0	1	0	1	0
g_n^+	1	2	3	4	5	6
g_n^{++}	1	1	1	1	1	1
g_n^{+-}	0	0	0	0	0	0

Table 1.6 Grundy numbers for MAG-NETISM

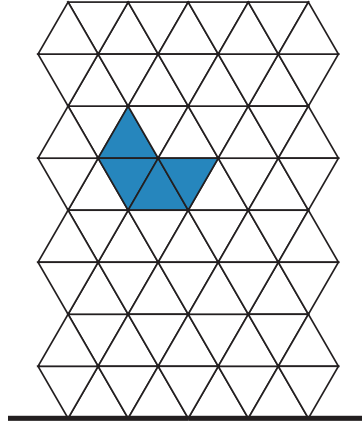


Figure 1.16 Example Position

The grid extends to the left and right as needed.

A move consists of transforming the tetramond into a new tetramond by moving one triangle downwards to a new position. Triangles cannot be moved below the floor line.

Below are all possible moves for the starting position to the right.

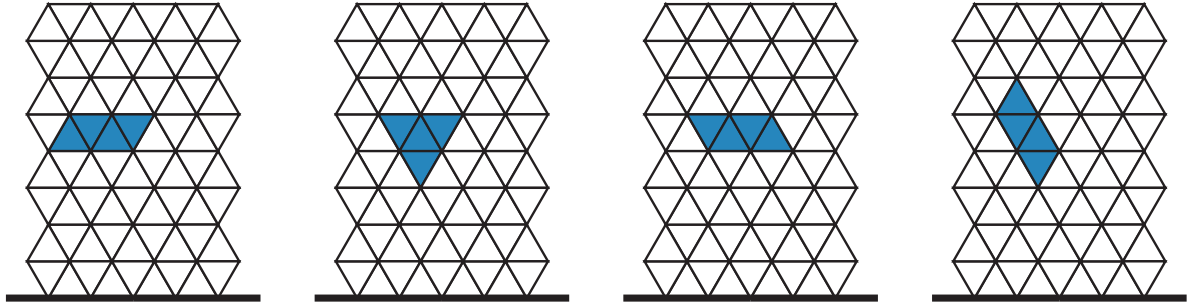


Figure 1.17 Options

Note that there are just three tetramonds, which can occur in different orientations.

The game ends when no move is possible anymore, and the player who moves last wins.

Let's denote a game position by a pair of numbers (s, h) where s indicates the shape of the tetramond as in [Figure 1.18](#), and h denotes the height of the shape above the ground. We didn't record tetramonds twice that were just symmetric with respect to a reflection at a vertical line, so there are eight different shapes to consider. The height h is measured as the distance from the lowest point of the tetramond to the ground line. Thus in [Figure 1.18](#), all shapes are at height 1. The game ends when position is $(1, 0)$ is reached, because every other position still allows a move.

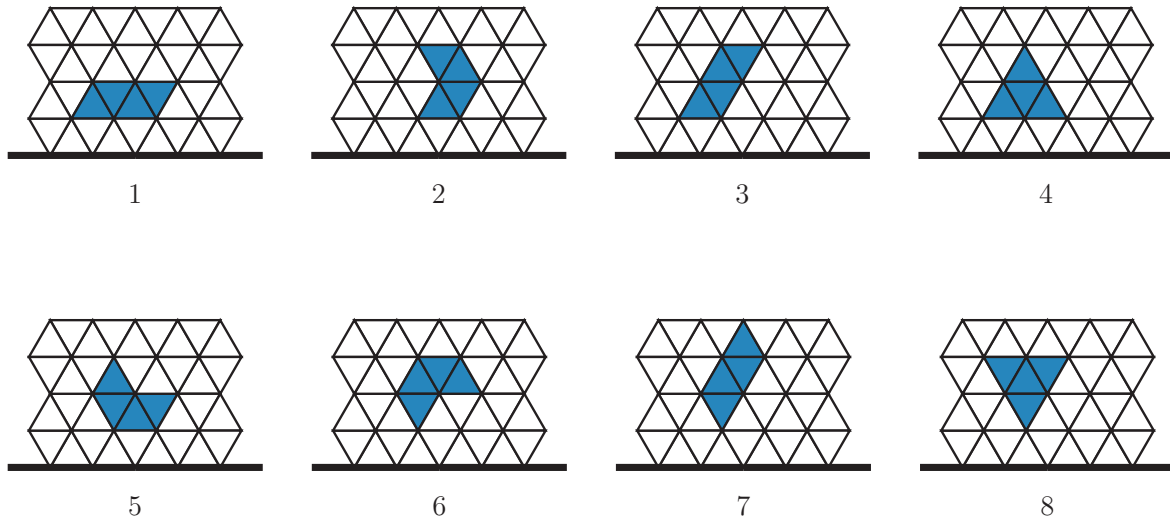


Figure 1.18 All eight tetriamond positions at height 1

Theorem 1.6.1 *The shapes 1, 2, and 3 are losing positions at any height, and all other shapes are winning positions.*

Proof. The following table lists all positions with their options, and the resulting Grundy numbers.

position	options	Grundy number
$(1, i)$	$(6, i - 1), (8, i - 1)$	0
$(2, i)$	$(4, i), (5, i), (7, i - 1)$	0
$(3, i)$	$(4, i), (5, i), (7, i - 1)$	0
$(4, i)$	$(1, i), (6, i - 1), (7, i - 1)$	3
$(5, i)$	$(1, i), (7, i - 1), (8, i - 1)$	3
$(6, i)$	$(2, i), (3, i)$	1
$(7, i)$	$(2, i), (3, i), (6, i), (8, i)$	2
$(8, i)$	$(2, i), (3, i)$	1

Table 1.7 Grundy numbers of TETRIAMOND

Remarkably, the Grundy number of a position depends only on the tetriamond but not its height. This can be explained as follows, resulting in a simple strategy:

Let's call a position even or odd, depending on the number of triangles at the lowest level of the tetriamond. The shapes 1, 2 and 3 are even, while the others are all odd.

From the table, we see that for every even tetriamond, the next move takes us to an odd tetriamond. On the other hand, for all odd tetriamonds, there is always an option that is even. Thus, the strategy is to always move to an even position. $\square$

SPRING CLEANING is played on a rectangular array of dirt pieces. A *sweep* consists of removing a single row or column of consecutive dirt pieces, as shown in Figure 1.19. A sweep must clean out as many dirt pieces as possible

in the direction of the sweep. A few illegal sweeps are shown in **Figure 1.20**. The two players take turns, doing one sweep at a time. The last player to sweep wins.

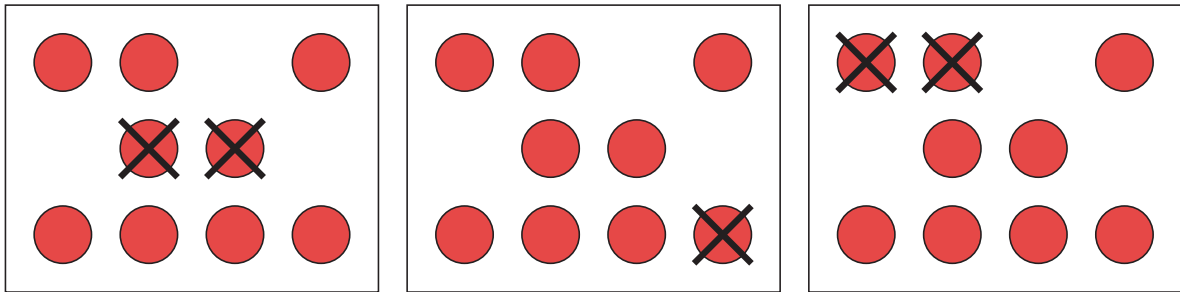


Figure 1.19 Legal Sweeps

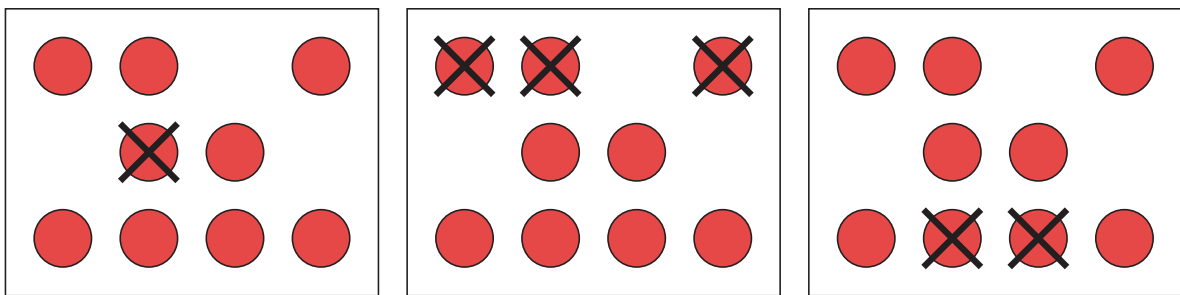


Figure 1.20 Illegal Sweeps

Exercise 1.6.2 Denote by $R_{a,b}$ a rectangular array of $a \times b$ dirt pieces. Compute the Grundy numbers $\mathcal{G}(R_{a,b})$ for $1 \leq a, b \leq 6$. Is there a pattern?

Solution:

$a \backslash b$	1	2	3	4	5
1	1	2	1	2	1
2	2	0	2	0	2
3	1	2	1	2	1
4	2	0	2	0	2
5	1	2	1	2	1

Table 1.8 $\mathcal{G}(R_{a,b})$

This leads to the general formula

$$\mathcal{G}(R_{a,b}) = \begin{cases} 0, & \text{if } a \text{ and } b \text{ are both even} \\ 1, & \text{if } a \text{ and } b \text{ are both odd} \\ 2, & \text{if } a \text{ and } b \text{ have different parity} \end{cases}$$

which can be proved by induction.

Exercise 1.6.3 Find at least one winning move for the position shown in **Figure 1.21**, and explain.

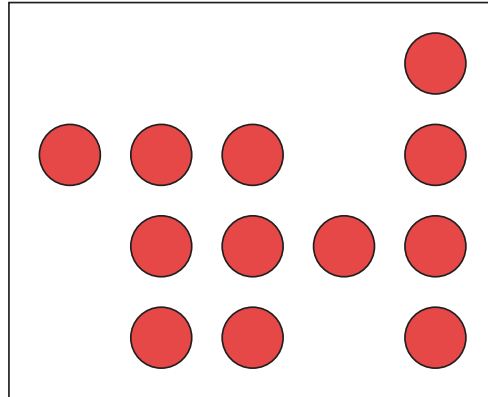


Figure 1.21 Find a winning move!

Solution: The move indicated in **Figure 1.22** breaks the game into the sum $R_{1,3} + R_{1,2} + R_{2,1} + R_{1,1}$ which has Grundy number $1 \oplus 3 \oplus 3 \oplus 1 = 0$, so this is indeed a winning move.

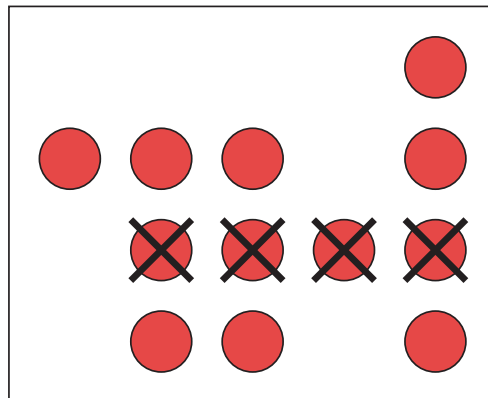


Figure 1.22 A winning move!

Exercise 1.6.4 Denote by $L_{a,b}$ an L-shaped region with a vertical and b horizontal dirt pieces. Thus $L_{a,1} = R_{a,1}$ and $L_{1,b} = R_{1,b}$. Compute the Grundy numbers $\mathcal{G}(L_{a,b})$ for $2 \leq a, b \leq 6$.

Solution: We compute by hand that

This leads to the general formula for $a + b > 3$:

$$\mathcal{G}(R_{a,b}) = \begin{cases} 0, & \text{if } a \text{ and } b \text{ have different parity} \\ 1, & \text{if } a \text{ and } b \text{ are both odd} \\ 3, & \text{if } a \text{ and } b \text{ are both even} \end{cases}$$

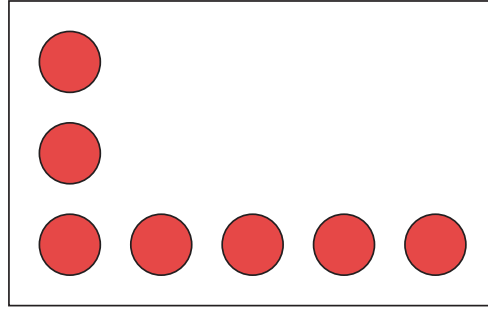


Figure 1.23 $L_{3,5}$

a\b	2	3	4	5	6
2	0	4	3	0	3
3	4	1	0	1	0
4	3	0	3	0	3
5	0	1	0	1	0
6	3	0	3	0	3

Table 1.9 $\mathcal{G}(L_{a,b})$

and the ‘exceptional’ values $L_{2,2} = 0$ and $L_{2,3} = L_{3,2} = 4$.

An easy to define class of impartial games are *subtraction games*:

subtraction
game

Definition 1.6.5 A subtraction game is determined by a set of permissible numbers $S = \{p_1, p_2, \dots\}$ and played with a collection of piles like NIM. A move consists of removing from one pile a permissible number of tokens.

Example 1.6.2

1. If $S = \mathbb{N}$, all numbers are permissible, and we are just playing NIM.
2. If $S = \{1, 2, \dots, p\}$ we are playing the game $N(p, n)$ from [Corollary 1.5.10](#).
3. Let $P = \{1, 3\}$. Computing the Grundy numbers for playing a single pile of size n reveals that

$$\mathcal{G}(n) = \begin{cases} 0, & \text{if } n \equiv 1(\text{mod } 4) \text{ or } n \equiv 3(\text{mod } 4) \\ 1, & \text{if } n \equiv 0(\text{mod } 4) \\ 2, & \text{if } n \equiv 2(\text{mod } 4) \end{cases}$$

This can be proven by induction.

Example 1.6.3 The computation of the Grundy numbers of a subtraction game can be streamlined by using a *Grundy Ruler*: In [Figure 1.24](#) the Grundy ruler connects its dot on the right with arrows to the fields 1, 3, and 4 positions to the left. To determine the Grundy number at the dot, we just need to compute the mex of the Grundy numbers at the arrows. We can then continue the computation by sliding the ruler to the next empty field on the right.

In all the examples so far, the Grundy numbers of a subtraction game have become periodic after a while.

n	0	1	2	3	4	5	6	7	8	9	10	11	12	13
$\mathcal{G}_n$	0	1	0	1	2	3	2	0						



Figure 1.24 The Grundy ruler for the subtraction game with $S = \{1, 3, 4\}$

Theorem 1.6.6 Let $S = \{s_1, \dots, s_n\}$ be a finite subtraction set. Denote by $\mathcal{G}_n$ the Grundy number of the corresponding subtraction game. Then $\mathcal{G}_n$ becomes eventually periodic, i.e. there is an integer $p > 0$ (the period) and an index N such that $\mathcal{G}_{n+p} = \mathcal{G}_n$ for all $n > N$.

Proof. Let $m = \max S$ be the largest number in the subtraction set. Any block of m consecutive Grundy numbers determines the rest of the sequence, because we can inductively find the following Grundy numbers by looking at at most m previous Grundy numbers.

If we therefore can find two identical blocks of length at least m in the sequence $\mathcal{G}_n$ of Grundy numbers, we are done: The elements after the first and second block must be identical. In particular, after the second block, there will appear a third, fourth, etc block.

Furthermore, the largest Grundy number that can occur is m , as a block of m numbers cannot comprise all numbers between and including 0 and m .

But there are only finitely many blocks of length m consisting of numbers less than $m + 1$, namely m^{m+1} . Therefore, a long enough sequence of Grundy numbers will contain two identical blocks of length m , which we had to establish.

Exercise 1.6.7 Let $S = \mathbb{N} - \{2, 3\}$, i.e. you can take from a pile any number of tokens but 2 or 3. Compute the first Grundy numbers until you discover a pattern. Prove the pattern using induction.

Solution: The first Grundy numbers are 0, 1, 0, 1, 2, 3, 2, 3, 4, 5, 4, 5, 6, 7, 6, 7, 8, 9, 8, 9.

Exercise 1.6.8 This game is played with a single pile of size n . If n is even, you are allowed to take 1 or 3 tokens, and when n is odd, you can take 1 or 2 tokens. Determine the Grundy value for piles of size 0 through 16 until you discover a pattern.

Solution: The first Grundy numbers are 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0, 2, 0, 1, 0.

Example 1.6.4 In the Rip van Winkle Puzzle, SAM LOYD introduces a game that is now called KAYLES ([3]). It is being played with n pins in a row. The players can knock out one or two consecutive pins with one shot. Their honor prevents them from missing the pins altogether. They are also skillful enough to always hit the pins they desire to knock out. The two players take turns, and the player who shoots down the last pin wins. As is, this game is an easy win for the first player, who can just knock out the one or two central pins, and from then on play symmetrically. However, in Loyd's problem, of the 13 consecutive pins, the second one is already knocked out, i.e., missing. The question is which pin the first player has to knock out to win.

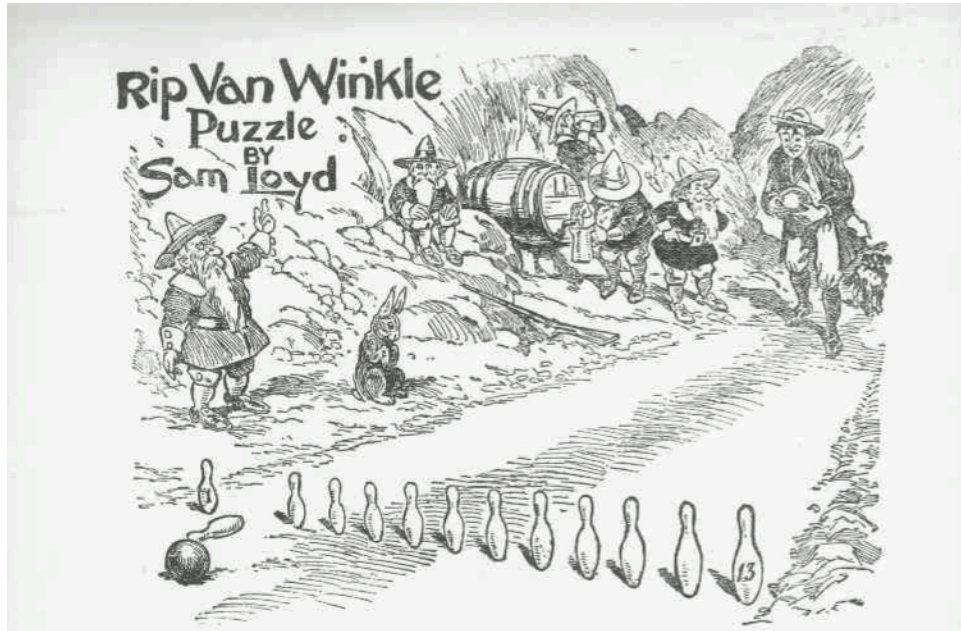


Figure 1.25 The Rip van Winkle puzzle

Exercise 1.6.9 For the game KAYLES,

1. Compute the Grundy numbers for all positions with up to 13 consecutive pins. For example, you should find that $\mathcal{G}(11) = 6$.
2. What are all the winning moves for the initial position with 13 pins?
3. Solve Sam Loyd's puzzle.

Solution:

1. Let K_n be a KAYLES position with n consecutive pins, and let $g_n = \mathcal{G}(K_n)$. Then $g_0 = 0$ and $g_1 = 1$. Note that any move breaks K_n either into a game sum $K_k + K_{n-k-1}$ (when knocking out one pin) or $K_k + K_{n-k-2}$ (when knocking out two pins) Thus,

$$g_n = \text{mex}(g_0 \oplus g_{n-1}, \dots, g_{n-1} \oplus g_0, g_0 \oplus g_{n-2}, \dots, g_{n-2} \oplus g_0)$$

This allows us to find

n	1	2	3	4	5	6	7	8	9	10	11	12	13
g_n	1	2	3	1	4	3	2	1	4	2	6	4	1

Table 1.10 Grundy numbers for KAYLES

2. The options for K_{13} are $K_{12}, K_1 \oplus K_{11}, K_2 \oplus K_{10} \dots$, for knocking out one pin and $K_{11}, K_1 \oplus K_{10}, K_2 \oplus K_9 \dots$. Using the table above, we compute

K	K_{12}	$K_1 \oplus K_{11}$	$K_2 \oplus K_{10}$	$K_3 \oplus K_9$	$K_4 \oplus K_8$	$K_5 \oplus K_7$	$K_6 \oplus K_6$
$\mathcal{G}(K)$	4	$1 \oplus 6 = 7$	$2 \oplus 2 = 0$	$3 \oplus 4 = 7$	$1 \oplus 1 = 0$	$4 \oplus 2 = 6$	$3 \oplus 3 = 0$
K	K_{11}	$K_1 \oplus K_{10}$	$K_2 \oplus K_9$	$K_3 \oplus K_8$	$K_4 \oplus K_7$	$K_5 \oplus K_6$	
$\mathcal{G}(K)$	6	$1 \oplus 2 = 3$	$2 \oplus 4 = 6$	$3 \oplus 1 = 2$	$1 \oplus 2 = 3$	$4 \oplus 3 = 7$	

Table 1.11 Grundy numbers for the options of K_{13}

Without (re-)computing the table, we could also have observed that a winning move must split K_{13} into $K_m + K_n$ so that $g_m = g_n$ and $m + n = 12$ or $m + n = 11$. Thus, the winning moves are those that knock out either pin at the positions 3, 5, 7, 9, 11.

- Knocking out the leftmost pin is no good, as $g_{11} = 6 \neq 0$. The remaining options split K_{11} into two piles of size m and n with $m + n = 10$ or $m + n = 9$. Such an option is a winning move if and only if $1 \oplus g_m \oplus g_n = 0$. Inspecting **Table 1.10** shows that we need to have $(m, n) = (3, 7)$ or $(m, n) = (7, 3)$. This means that in Loyd's puzzle, we need to knock out either pin 6 or pin 10.

1.7 COIN-TURNING GAMES

In a coin-turning game, coins are arranged in a row, showing heads and tails. A move consists of turning one coin from tails to heads, and a number of coins to the right of it either way. Which coins may be turned depends on the specific game. These games have been invented and analyzed by HENDRIK WILLEM LENSTRA.

We begin with TURNING TURTLES. It is being played with several turtles in a row, either on their feet or upside down. A move consists of flipping over up to two turtles: The leftmost of the two turtles must be turned from upright to upside down. A second turtle to the right may be turned either way. We label the locations of the turtles from right to left with 1, 2, 3, etc.

To analyze the game, we will leave the turtles in peace and use instead the digits 1 and 0 to designate an upright and an upside-down turtle, respectively. This way, any position becomes a binary number. Any move will decrease that number, as the rules impose that the leftmost change is a change from a 1 to a 0. Therefore, any such game will end.

This game is yet another form of NIM. To see this, we represent a digit 1 at position i by a NIM pile of size i . We now claim that the possible options for TURNING TURTLES correspond to the options for the associated NIM game.

To see this, let's consider all the possible moves in a TURNING TURTLES position. There are three possibilities:

1. Change a digit 1 at position i into a 0, and leave all other digits alone. This corresponds to completely eliminating the pile of size i in NIM.
2. Change a digit 1 at position i into a 0, and also change a digit 1 at position $j < i$ to 0. This corresponds to completely replacing the pile of size i by a pile of size j .
3. Change a digit 1 at position i into a 0, and change a digit 0 at position $j < i$ to 1. This corresponds to eliminating the piles of size i and j . This is not a legal NIM move, but it is equivalent to replacing the pile of size i with a second pile of size j , as two equally sized NIM piles cancel.

The same argument allows us to reduce all coin-turning games to the analysis of special cases: Denote by T_n the position that has turtles upside down in locations 1 through $n - 1$, and a single turtle being upright at location n . See **Figure 1.26** for an example. The position T_0 stands for the empty board without a move. Then we claim:

Theorem 1.7.1 *Suppose a position T in a coin turning game has a digit 1 at positions $k_1, \dots, k_n$. Then $T = T_{k_1} + \dots + T_{k_n}$, possibly up to adding a game with Grundy number 0.*

Proof. We again show that the options in T correspond naturally to the options in the sum $T_{k_1} + \dots + T_{k_n}$. A move in T consists of picking a location i_1 with a digit 1, a number of locations $i_2, i_3, \dots$, then changing the 1 at i_1 to a 0, and flipping the digits at the other chosen positions. We translate this into a move in the sum $T_{k_1} + \dots + T_{k_n}$ by replacing the summand T_{i_1} with new summands $T_{i_2}, T_{i_3}, \dots$. If we get two equal summands this way, we just remove them, as their Grundy number is 0. $\square$

The following game is called MOCK TURTLE and allows to turn over three turtles, coins, or better digits: The leftmost digit must again turn from 1 to 0, while the other two can be changed either way.

To analyze this game, we will need to divide the natural numbers into two classes, called *odious* and *evil*:

Definition 1.7.2 *A non-negative integer is called odious if the number of 1s in its binary expansion is odd; otherwise, it is called evil.*

The first odious numbers are $O_n = 1, 2, 4, 7, 8, 11, 13, 14, 16, 19, \dots$

Exercise 1.7.3 Show that the NIM-sum of two odious or two evil numbers is evil, while the NIM-sum of an odious and an evil number is odious.

We now claim

Theorem 1.7.4 For MOCK TURTLES, we have $\mathcal{G}(T_n) = O_n$.

Proof. We will prove this by induction over n . For $n = 1$, clearly $\mathcal{G}(T_1) = O_1 = 1$. Now let's assume the claim has been verified for all integers up to $n - 1$. To compute $\mathcal{G}(T_n)$, we first determine all possible options and their Grundy numbers: One option is to just flip to 1 at location n , reaching T_0 , which has Grundy number 0. Another set of options changes the 1 at location n to 0 and changes a 0 at any location j between 1 and $n - 1$ to 1. This leads to a position T_j whose Grundy numbers are all odious numbers $O_1, \dots, O_{n-1}$ by induction hypothesis. Finally, we can change the 1 at location n to 0 and change the zeroes at two locations $j < k$ between 1 and $n - 1$ to 1. This leads to a position $T_j + T_k$ which has Grundy number $O_j \oplus O_k$, again by induction hypothesis. We now need to show that the mex of all these Grundy numbers is equal to O_n . Because the sum of two odious numbers is evil, the first odious number not in the set is O_n . We thus only need to show that all evil numbers that are smaller than O_n are either 0 or the NIM sum of two smaller odious numbers. If an evil number is non-zero, its binary representation has at least two 1s and can therefore be written as the NIM sum of two different smaller odious numbers. $\square$

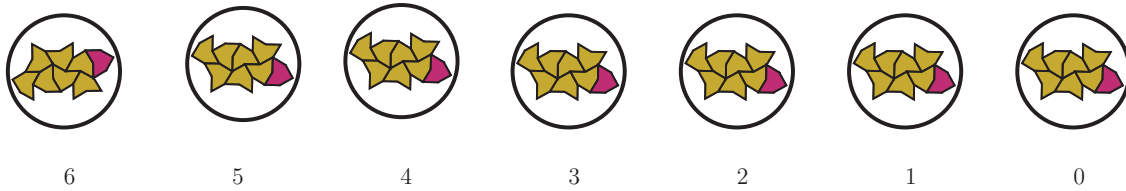
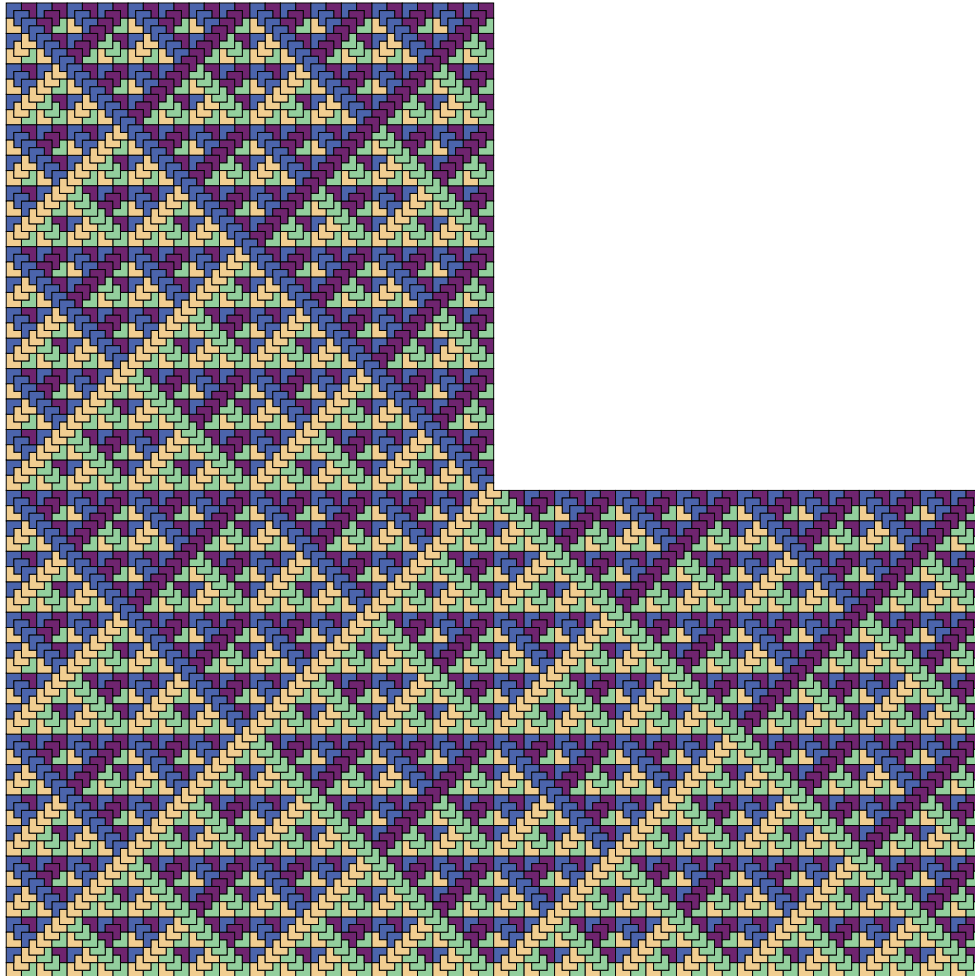


Figure 1.26 Position T_6

2 Polyforms

*You're old enough now, Duncan, to understand this game...
though it's very much more than a game.*

ARTHUR C. CLARKE, *Imperial Earth*



polyform

Definition 2.5 *A polyform is a shape that consists of several congruent shapes glued together edge-to-edge.*

There are many different kinds of polyforms. Most frequently used are *polyominoes*, *polyiamonds*, and *polyhexes*. Depending on the number of shapes used in a polyform, they are called *monoforms*, *duoforms*, *triforms*, *tetraforms*, or similarly, using Latin or Greek prefixes for the numbers.

Polyforms will provide us with a wealth of interesting shapes and puzzles. We will classify them, use them to tile regions in the plane, and use coloring arguments to show that this is sometimes not possible.

2.1 POLYOMINOES

pentomino

In 1953, SOLOMON W. GOLOMB [4] invented pentominoes: Twelve shapes that can be assembled by joining unit squares edge-to-edge. It turned out that this small set of shapes was well-suited for a large variety of puzzles. We will study some of them.

<i>name</i>	<i>number</i>
Monomino	1
Domino	1
Tromino	2
Tetromino	5
Pentominos	12
Hexominos	35

Table 2.1 Number of n -ominoes for small n

polyomino

More generally, *polyominoes* arise by gluing together squares of the same size edge to edge. We will consider two polyominoes as equal if they are congruent by translation, rotation, or reflection. The table lists the names together with the number of different kinds. Computer programs have been written to enumerate all n -ominoes for a given n , but no general formula is known.

In [Figure 2.1](#) we have listed all polyominoes with up to five squares. They have been given names, using letters of the alphabet that look more or less like the polyomino. Note again that we don't distinguish polyominoes that are congruent by reflection.

Our primary concern with polyominoes is the following fundamental

Problem 2.1.1 *Can one tile a given region in the plane with a given set of polyominoes?*

Here we mean by to *tile* that the pieces don't overlap and leave no gaps.

For instance, it is easy to tile the 8×8 chessboard with 32 dominoes, and impossible to tile any region consisting of an odd number of squares with dominoes.

Now, any single problem can be solved with a lot of patience by trial and error, which will usually take a long time. The time it takes a computer program to tile a general region with n squares grows exponentially in n , for all known algorithms.

In many cases, we can use clever insight to either find a tiling or claim that this is impossible, and not only for a single region, but for a whole class of regions at once. Thus, we are not so much looking for an answer to the problem above in particular cases, but rather for general insight.

The following is an example of that and will introduce us to a rather powerful method.

Exercise 2.1.2 *Remove two diagonally opposite corners from an 8×8 chessboard as in [Figure 2.2](#). Can you tile the remaining board with 31 dominoes? Every square must be covered, and no tiles may overlap.*

The solution uses a *coloring argument*: Using the standard black and white coloring of the chessboard, each domino will cover one black and one white square, no matter how it is placed. Thus, in any tiling by dominoes, the same number of black and white squares will be covered. For our amputated chessboard, we have removed two squares of the same color, making it impossible to tile.

One could, of course, solve such a problem by trial and error. However, having a general method of solution at hand allows one to solve many problems simultaneously. For instance, now we know that removing opposite corners on *any* $n \times n$ board will render the board impossible to tile by dominoes.

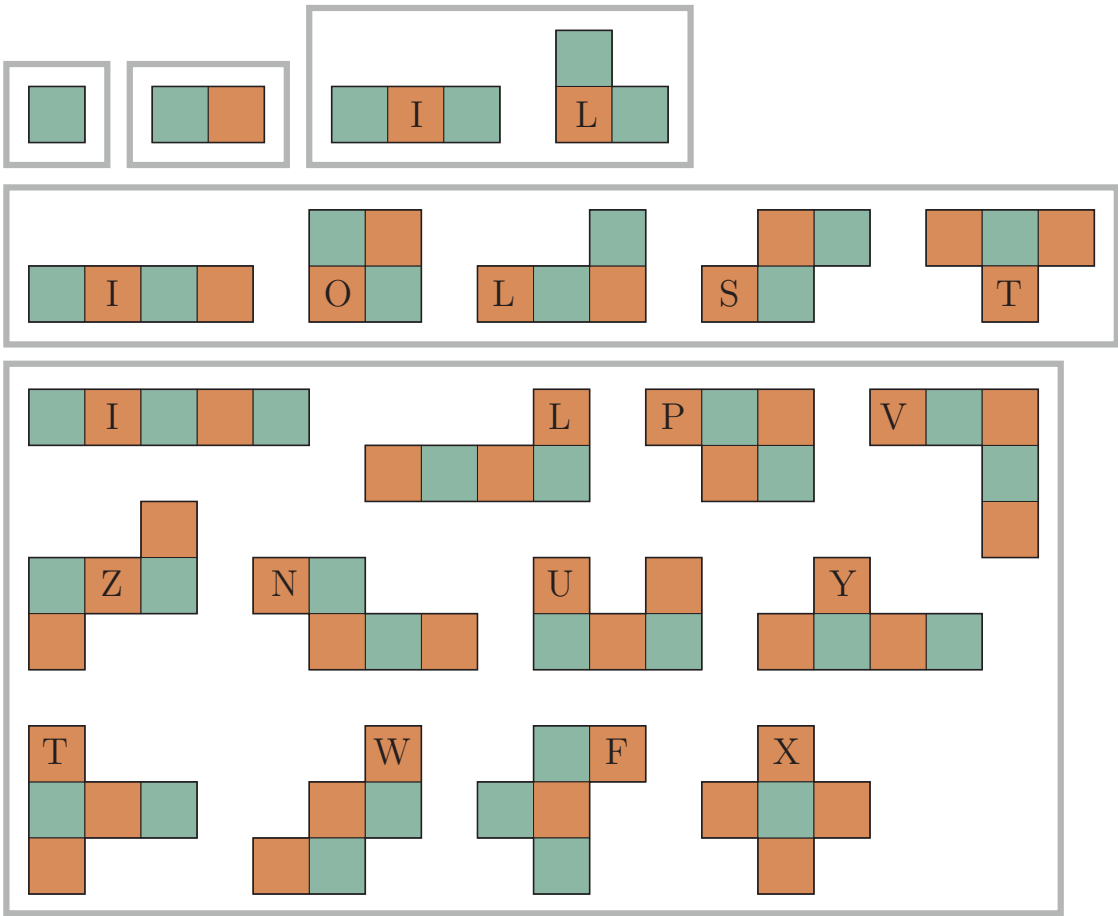


Figure 2.1 The mono-, do-, tro-, tetra-, and pentominoes, up to symmetries

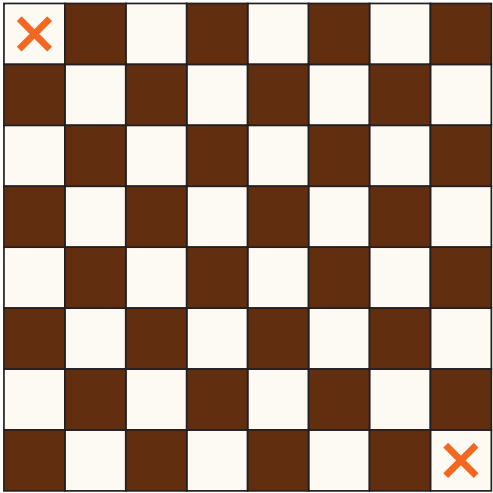


Figure 2.2 Tile me with dominoes

The same kind of argument can be applied in many other cases.

Exercise 2.1.3 Can one tile a 4×5 rectangle with the five tetrominoes, using each tetromino once?

Solution: This is impossible: Color the 4×5 -rectangle with black and white squares in a chessboard fashion, using ten squares of each color. The S-, O-, I-, and L-tetromino will each cover two black and two white squares, but the T-tetromino will cover three black squares and one white square, or vice versa.

Very often, one can refine this kind of argument by using other colorings of the square tiling. Here is one that uses three colors:

Exercise 2.1.4 You are allowed to remove a single square from the 8×8 chessboard in order to tile it with I-trominoes. Which square can you choose?

Observe that the I-tromino will cover one square of each color in the 3-coloring of the 8×8 square in **Figure 2.3**.

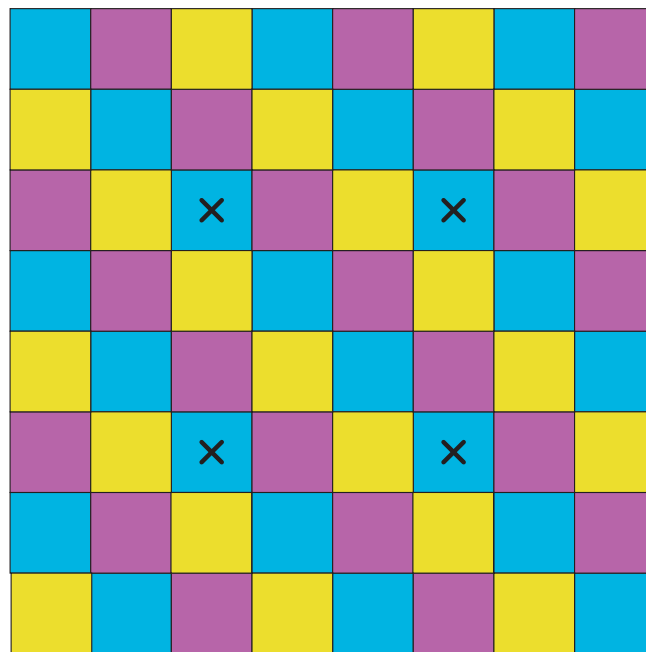


Figure 2.3 Tiling with straight trominoes

This coloring uses 21 yellow and purple squares, but 22 blue squares. Therefore, we clearly need to remove a blue square. But that is not all: While the square board is symmetric with respect to 90° rotations, the coloring isn't. Therefore, we can use a second coloring by rotating the board by 90° . The square to be removed must then also belong to those that are colored blue in the rotated coloring. This leaves us with the four marked squares as squares that we still have a chance to remove. And indeed, it is possible to tile the square with any of these marked squares removed. To see this, first tile the perimeter outside the inner 4×4 square with I-trominoes, and then proceed to the inner square.

Definition 2.1.5 Given a coloring of the infinite chessboard, we say that a polyomino is balanced if it covers the same number of squares of each color that is used in the coloring, no matter how it is placed.

For instance, the domino is balanced with respect to the black and white coloring, while the I-tromino is balanced with respect to the blue-yellow-purple coloring.

Our coloring argument shows:

Theorem 2.1.6 *A set of balanced polyominoes cannot tile an unbalanced region.*

There are variations of this argument, exploiting *how* unbalanced the situation is:

Example 2.1.1 Can one tile the chessboard with one T-tetromino and 15 S-tetrominoes? We know that the S-tetromino is balanced, so 15 of them will cover 15 white and 15 black squares. There remain two white and two black squares to be tiled, but the T-tetromino is unbalanced, so this is impossible.

Exercise 2.1.7 *Determine for all hexominoes which are balanced with respect to the black and white coloring.*

Exercise 2.1.8 *Determine for all hexominoes which are balanced with respect to the blue-yellow-purple coloring.*

Exercise 2.1.9 *Can the D-hexomino  tile a 9×10 -rectangle? Hint: Find a coloring with two colors so that the D-hexomino is balanced, no matter how it is placed, but the rectangle is balanced.*

Solution: Color the rectangle as shown in **Figure 2.4**. Then a D-hexomino covers both colors equally often. But the 10×0 -rectangle covers two more red tiles than blue tiles.

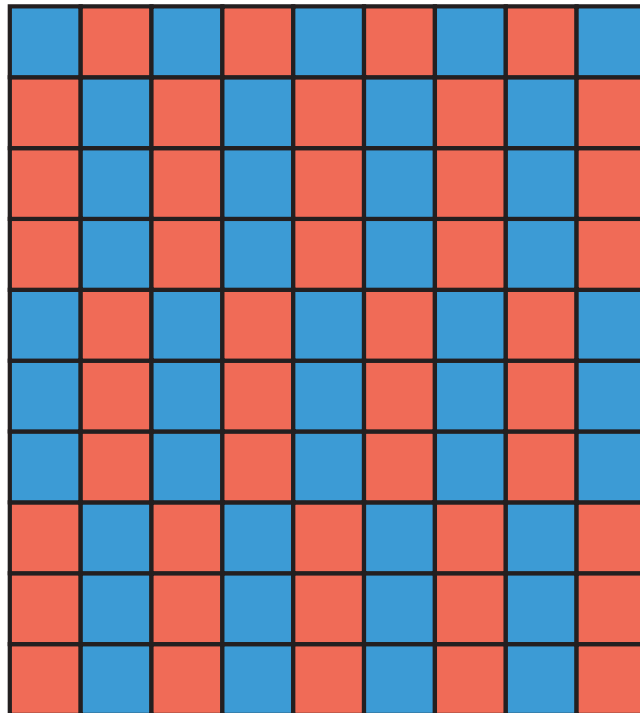


Figure 2.4 Tiling with D-hexominoes

Exercise 2.1.10 Can the 9×12 rectangle be tiled with the -hexomino? We allow for rotations and reflections.

Solution:

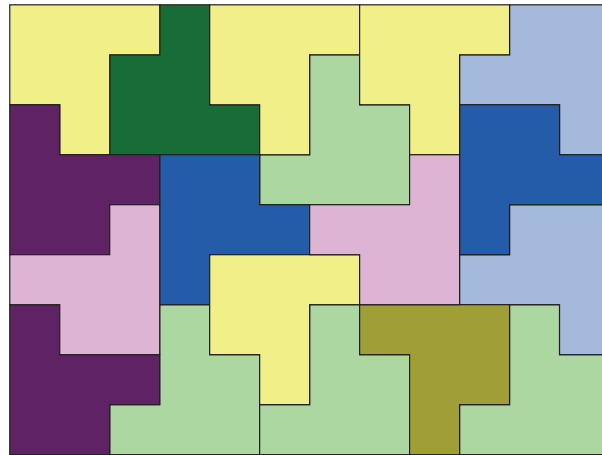


Figure 2.5 Tiling the 9×12 rectangle with the -hexomino

There is another useful coloring with two colors that is shown in **Figure 2.6** which we will explore in the following exercises.

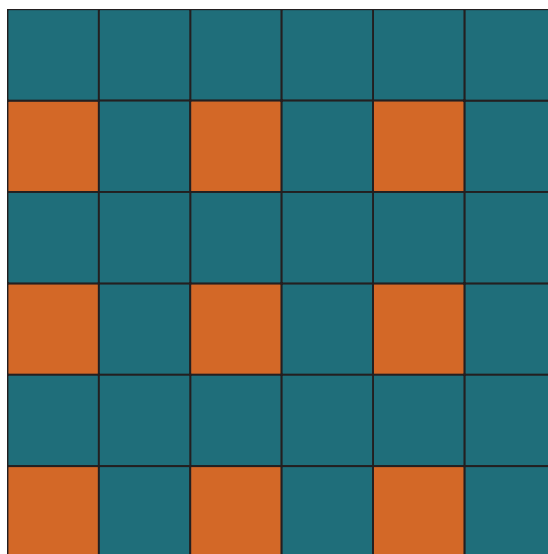


Figure 2.6 Another 2-coloring of the square tiling

Exercise 2.1.11 Which of the following rectangles can be placed on the 2-colored square tiling of Figure 2.6 so that the number of orange squares is odd?

1. 6×8
2. 6×10
3. 7×7

Solution: According to the following exercise, no even edge may be divisible by 4. Thus, only the 6×10 and the 7×7 rectangle can be placed so that they cover an odd number of orange squares.

We now generalize this:

Exercise 2.1.12 For which dimensions can an $m \times n$ -rectangle be placed on the 2-colored square tiling of Figure 2.6 so that the number of orange squares is odd?

Solution: We make a case distinction:

If m and n are both even, then the number of orange squares is $\frac{1}{4}mn$, no matter how the rectangle is placed. In order to have a total number of odd orange squares, this means that we need to have $m = 2a + 2$ and $n = 4b + 2$.

If m is even and n is odd, we can slide the rectangle up and down to change the number of orange squares by $\frac{1}{2}m$. Sliding the rectangle left or right doesn't change the number of orange squares. Thus we either have $\frac{1}{4}m(n + 1)$ or $\frac{1}{4}m(n - 1)$ orange squares. To make one of these numbers odd, we need that the even factor m and $n \pm 1$ are not divisible by 4. We can achieve this for $n \pm 1$ by choosing the right sign. Thus we can arrange for an odd number of squares if and only if $m = 4a + 2$.

The case that m is odd and n is even is similar.

Finally, if both m and n are odd, we have four possibilities for all sign choices in $\frac{1}{4}(m \pm 1)(n \pm 1)$. This can always be made even. In summary: A rectangle can be placed so as to have an odd number of orange squares if and only if the even edges are not divisible by 4.

To use this to show that certain polyominoes do not tile some rectangles, we need to find those polyominoes that always cover the same number of orange squares, counting modulo 2:

Exercise 2.1.13 Which hexominoes cover an odd number of orange squares in the tiling of Figure 2.6, no matter how they are placed?

Solution: These are  and .

Exercise 2.1.14 Which hexominoes cover an even number of orange squares in the tiling of Figure 2.6, no matter how they are placed?

Solution: There are , , , and .

Here is an example of how to apply this:

Exercise 2.1.15 Show that neither the -hexomino nor the -hexomino can tile a the 18×20 rectangle.

Solution: To tile the 18×20 rectangle, we need 60 hexominoes, which is an even. The -hexomino covers an even number of orange squares, so 60 of them will do so as well. The -hexomino covers an odd number of orange squares, so 60 of them will cover an even number of orange squares again. But the 18×20 rectangle can be placed so that it covers an odd number of orange squares.

Now we generalize again:

Exercise 2.1.16 Show that if m and n are even but not divisible by 4, then an $m \times n$ rectangle cannot be tiled by any one of the hexominoes you found in the exercises above.

Solution: The argument is the same as in the previous exercise.

Exercise 2.1.17 Can one tile the 6×6 -square with L -tetrominos? Hint: Put weight 0 on all squares (x, y) with x even and weight 1 on all squares with x odd.

Solution: Any placement of the L -tetromino will have odd weight, and an odd number of them will still have odd weight. But the 6×6 -square has an even weight.

It is sometimes more practical to use colorings by numbers (*weights*), as the following example shows. Label horizontal rows of squares alternately by the numbers 1 and 5 as in [Figure 2.7](#). The *weight* of a set of squares is defined as the sum of all numbers in the squares. Then, an L -tetromino will cover a region whose weight is either 8 or 16, depending on how it is placed. Therefore, any region covered by L -tetrominoes must have a weight divisible by 8.

Now suppose we have a rectangle of dimensions $a \times b$. In order to be able to tile it with the L -tetromino, we need that ab is divisible by 4, just by area consideration. Let's assume, for instance, that b is even. Then the rectangle will cover b pairs of rows that have weight a and $3a$ each. Thus, the total weight of the rectangle will be $3ab$, which must be divisible by 8, as we have seen. This shows:

Theorem 2.1.18 Any rectangle whose area is not divisible by 8 cannot be tiled by L -tetrominoes.

Exercise 2.1.19 Find a tiling of a 3×8 rectangle with L -tetrominoes.

Solution:

Exercise 2.1.20 Find a tiling of a 5×8 rectangle with L -tetrominoes that uses only the L version, but not its mirror image.

Solution:

We now turn to “positive” results that state that certain shapes *can* be tiled.

We have learned that a rectangular chessboard cannot be tiled with dominoes if two squares of the same color have been removed. Surprisingly, however, we have

1	1	1	1	1	1	1	1
5	5	5	5	5	5	5	5
1	1	1	1	1	1	1	1
5	5	5	5	5	5	5	5
1	1	1	1	1	1	1	1
5	5	5	5	5	5	5	5
1	1	1	1	1	1	1	1
5	5	5	5	5	5	5	5

Figure 2.7 Values of L-tetrominoes

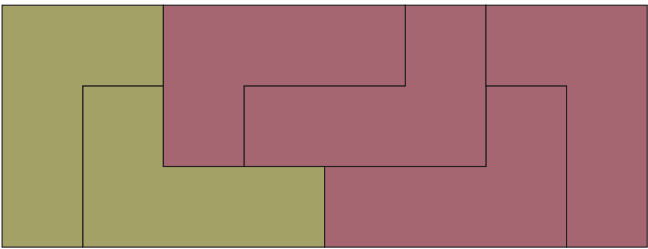


Figure 2.8 Tiling of a 3 × 8 rectangle with L-tetrominoes

Theorem 2.1.21 *The chessboard with any pair of squares of different colors removed can be tiled with dominoes.*

Proof. To prove a statement like this by trial and error would take a long time and would not be very illuminating. Fortunately, RALPH GOMORY has found a clever argument (see [5]).

Hamiltonian
cycle

We draw a closed circuit, a so-called *Hamiltonian cycle*, on the chessboard. There are many possibilities for that, and it doesn't matter how we do it. We can clearly tile the entire board with dominoes so that they follow the path, and thus so in precisely two different ways, depending on how we start. Now, remove two differently colored squares from the board. This disconnects the cycle into two paths. Each of them begins next to the removed square and ends next to the other removed square. Thus, the beginning and ending squares of each path have different colors as well, and the paths have even length. Note that if the two removed squares are neighbors, one of the two paths does not exist, i.e., it has length 0.

Therefore, we can tile both paths of the broken cycle with dominoes as desired. □

Exercise 2.1.22 *Show that Gomery's argument works for all rectangular boards with one edge of even length.*

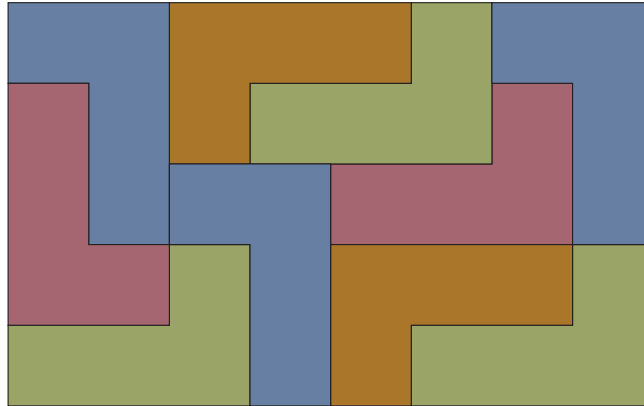


Figure 2.9 Tiling of a 5×8 rectangle with L -tetrominoes

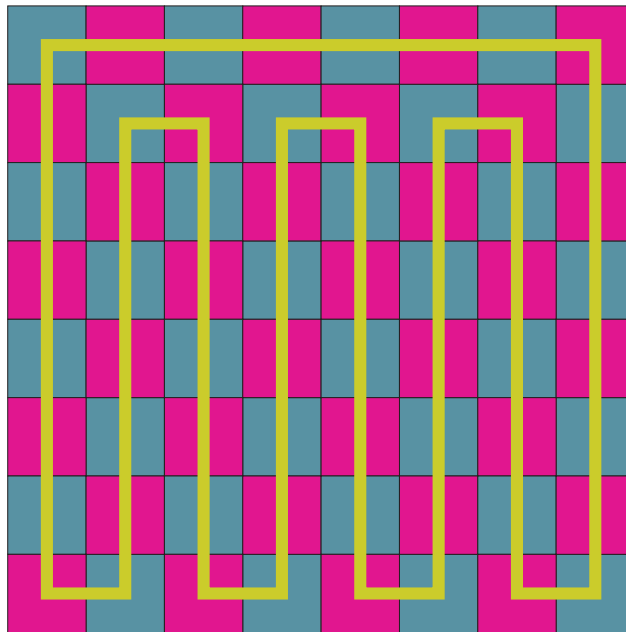


Figure 2.10 Gomory's Argument

Solution: The shape of the path in Figure 2.10 can be generalized to any board with one edge of even edge length.

Exercise 2.1.23 Let $n > m$ be positive integers so that $n - m$ is even. Remove from an $n \times n$ square and $m \times m$ square. Show that the resulting holey square has a Hamiltonian cycle.

Exercise 2.1.24 Suppose we have a rectangular board with both sides of odd length. Then we have one more square of one color than of the other color. Let's say there is one more white square. Modify Gomory's argument to show that any such board with any white tile removed can be covered with dominoes.

Solution: In this case, it suffices to draw an unclosed circuit that covers all squares of the board. To do so, you can use the same drawing as in Figure 2.10 with the top edge removed, for instance. Then the same argument works: The

path covers an odd number of squares and has thus white squares at both ends. Removing any white square will cut the path into two pieces, both of even length, each of which can be tiled by dominoes.

rep-tile

The next tilability argument makes use of the fact that the L -tromino is a *rep-tile*, i.e., a tile that tiles a larger copy of itself. We will discuss rep-tiles in more detail later; for now, the following will suffice:

Lemma 2.1.25 *Let n be a power of 2. Then the L -tromino tiles any L -tromino of size n .*

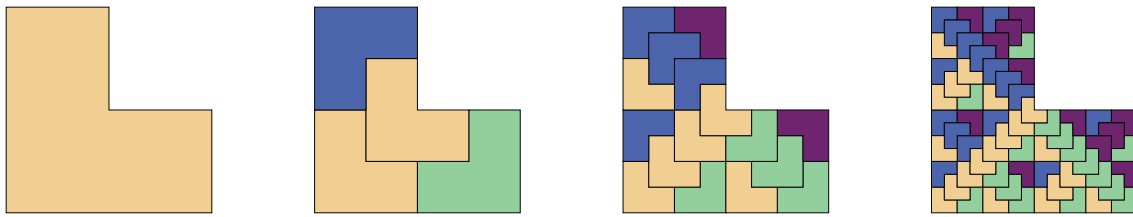


Figure 2.11 Successive subdivisions of the L -tromino into four smaller L -trominoes

Proof. Denote a tromino of size n by L_n . From Figure ?? we see L_2 can be tiled with four copies of L_1 . Now let n be a larger power of 2. Using the same figure at a larger scale, we can tile L_n with four copies of $L_{n/2}$. As $n/2$ is still a power of 2, we can continue tiling each $L_{n/2}$ with four copies of $L_{n/4}$, and thus obtain a tiling of L_n with 16 copies of $L_{n/4}$. We continue this way until we have tiled all L_2 -trominos with L_1 -trominos, and are done. $\square$

Proposition 2.1.26 *Let n be a power of 2. Then any $n \times n$ -board with any single square removed can be tiled by L -trominoes.*

Proof. Divide the given square Q_n of size n into four squares $Q_{n/2}$ of size $n/2$. As we have removed one square from Q_n , three of the smaller squares are complete and form an L_n -tromino, and the fourth square has just one small square missing. By the Lemma, we can tile the large L_n -tromino with L_1 trominos. Hence, it remains to tile a square of size $n/2$ with L -trominos. We repeat this argument until we end up with a square of size 2 with one square missing. This is, however, just an L_1 , and we are done. $\square$

Exercise 2.1.27 *Draw an 8×8 chessboard and remove a square of your choice. First, tile the board with L -trominos as in the proof. This will provide a single solution. Then, find a second, different solution.*

Solution: Two L -trominos fit together to tile a 2×3 -rectangle.

We first assume that the missing square is anywhere except in the central 4×4 -subsquare. Eight of the 2×3 -rectangles easily tile a solid 6×8 rectangle, and two more can be used to tile the remaining 2×8 -rectangle except for a 2×2 -square, which can be arranged to cover the missing square.

If the missing square is in the central 4×4 -subsquare, first tile that defective square as in the theorem, and then arrange eight 2×3 rectangles around the square.

Exercise 2.1.28 *Discuss which other defective rectangles can be tiled with L -trominoes.*

Exercise 2.1.29 An L-tromino comes in four different flavors: Depending on which square is missing from the 2-square, we call an L-tromino a NE-, NW-, SW-, or SE-tromino. Now take an 8×8 square, remove one of the squares of your choice, and tile the rest of the square with trominoes. Count how many trominoes of each flavor are needed.

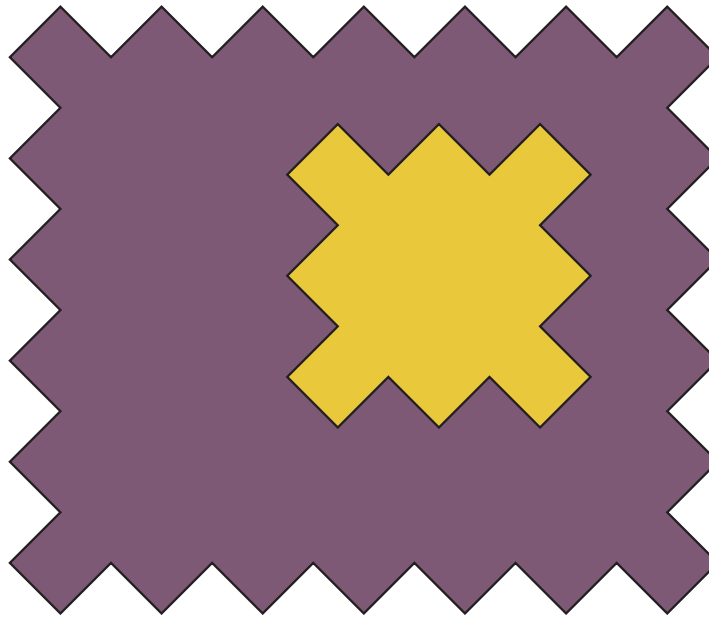


Figure 2.12 A ragged 3×3 square within a ragged 6×7 -rectangle

ragged
rectangle

Definition 2.1.30 A ragged rectangle of dimensions $m \times n$ is obtained from an $m \times n$ -rectangle by subdividing it into mn squares, centrally shrinking all squares slightly with respect to their respective centers, rotating them by 45° so that their vertices meet, and filling the gaps with $(m - 1) \times (n - 1)$ rotated squares.

Exploration 2.1.2 We are going to tile ragged rectangles with trominoes. In a ragged rectangle, the sides zigzag diagonally as in the figure, which shows a ragged rectangle of dimensions 6×7 . Note that the boundary changes directions at every unit step.

1. What is the area of a ragged rectangle of dimensions $a \times b$? Find a formula and test it on examples. Explain why you think the formula is correct.
2. Can one tile a ragged rectangle with dominoes?
3. For which a and b is the area of a ragged rectangle of dimension $a \times b$ divisible by 3?
4. Can you tile ragged 6×7 -rectangle with L-trominoes?
5. Can you tile the ragged 3×10 -rectangle with L-trominoes?
6. Can you tile the ragged 4×9 -rectangle with L-trominoes?
7. Which ragged rectangles do you think can be tiled with L-trominoes?

8. L-trominoes come in four different orientations. Can you tile a ragged rectangle with L-trominoes, using only three of the four orientations?
9. Investigate tilings of ragged rectangles by copies of a single tetromino.

Solution:

1. To compute the area of a ragged rectangle, color the squares alternately dark and light as in **Figure 2.13**. Then the outer set of dark squares is arranged in an array of $a \times b$ squares, while the inner light squares come in an array of $(a - 1) \cdot (b - 1)$ squares. Hence, we have a total of $ab + (a - 1)(b - 1)$ squares.

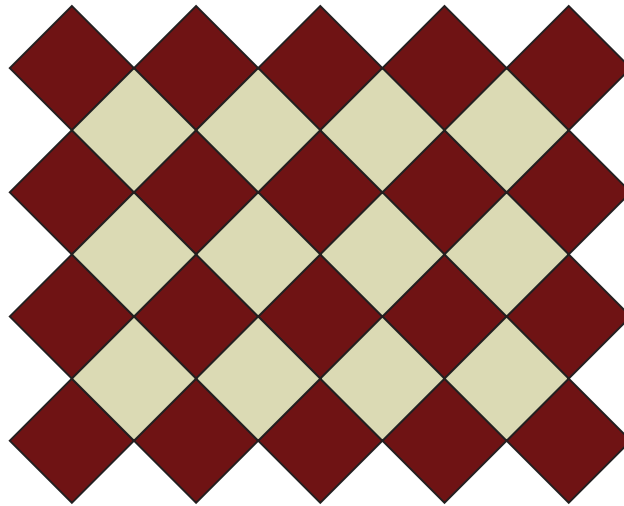


Figure 2.13 2-coloring of ragged rectangles

2. To test divisibility, we compute the area modulo 3, using the formula above, in the following table:

$a \backslash b$	0	1	2
0	1	0	2
1	0	1	2
2	2	2	2

This shows that the area of a ragged rectangle is divisible by 3 if and only if one dimension is divisible by 3 and the other has a remainder of 1 upon division by 3.

3. In order to tile a ragged rectangle by L-trominoes, we certainly need its area to be divisible by 3. It turns out that this is also sufficient (provided that the edge lengths are greater than 1). To see this, we proceed in several steps.

Figure 2.14 shows how to tile a ragged rectangle with dimensions 3×7 . By inserting more L-trominoes in blocks of five at the right end, we can tile this way any rectangle of dimensions $3 \times (3n + 1)$ with n at least 1.

Figure 2.15 shows how to tile a ragged rectangle with dimensions 4×6 . By inserting more L-trominoes in blocks of 7 at the right end, we can tile this way any rectangle of dimensions $4 \times (3n)$ with n at least 1.

Next, we put frames like the one in **Figure 2.16** around existing ragged rectangles. Such frames can be built to surround any ragged rectangle of dimensions $a \times b$ as long as $a, b \geq 1$, and they enlarge the rectangle to one with dimensions $(a + 3) \times (b + 3)$.

Figure 2.14 Tiling ragged $3 \times (3n + 1)$ rectanglesFigure 2.15 Tiling ragged $4 \times (3n)$ rectangles

These three constructions are sufficient to tile any ragged rectangle with dimensions $a \times b$ where a is divisible by 3 and $b \equiv 1 \pmod{3}$ (or vice versa). To do so, we first successively subtract 3 from both dimensions until the smaller dimension reaches 3 or 4, whichever happens first. Using one of the first two constructions, we tile that ragged rectangle with L-trominoes. Finally, we use the frame construction to enlarge that rectangle until we reach the original dimensions.

4. It is impossible to tile any ragged rectangle using only three of the four possible orientations of the L-tromino. To see this, we argue by contradiction. Let's assume that the down-pointing L-tromino is 'forbidden'. Then, to fill the top right corner of a ragged rectangle, we need to use the left-pointing L-tromino. Continuing to fill in the top edge leftwards, the exclusion of the down L-tromino forces us to use left-pointing L-trominoes, as in the image of the frame construction in Figure 2.16. We reach a contradiction at the top left edge, which cannot be tiled anymore.
5. Coloring the ragged rectangle black and white as in the proof of the area formula, we see that we have ab black squares but only $(a - 1)(b - 1)$ white squares. Hence, ragged rectangles are always unbalanced and cannot be tiled with the balanced tetrominoes. This leaves the T-tetromino as a candidate. Figure 2.17 shows how four 4×5 ragged rectangles, all tiled with T-tetrominoes, fit together to a larger 8×9 ragged rectangle.

Exploration 2.1.3 A *half-square* of edge length a is obtained from a square of size $a \times a$ by taking all unit squares that lie below or on a diagonal. For example, Figure 2.18 shows a half-square of edge length 8. We will explore the possibilities of tiling half-squares with various polyominoes.

1. For which a is the area of a half-square divisible by 3?
2. For small a , find which half squares can be tiled with L-trominoes.

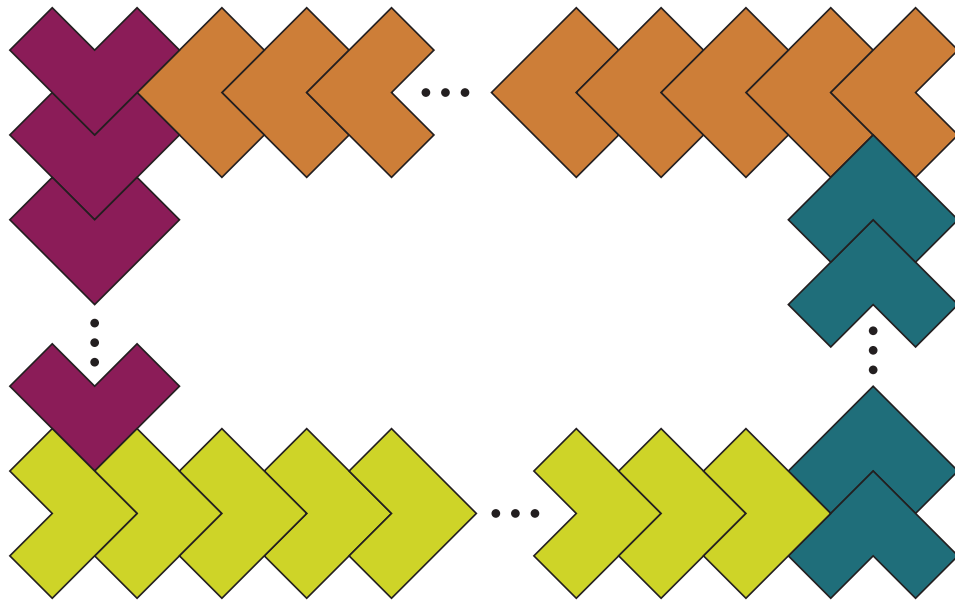


Figure 2.16 Putting ragged rectangles into frames

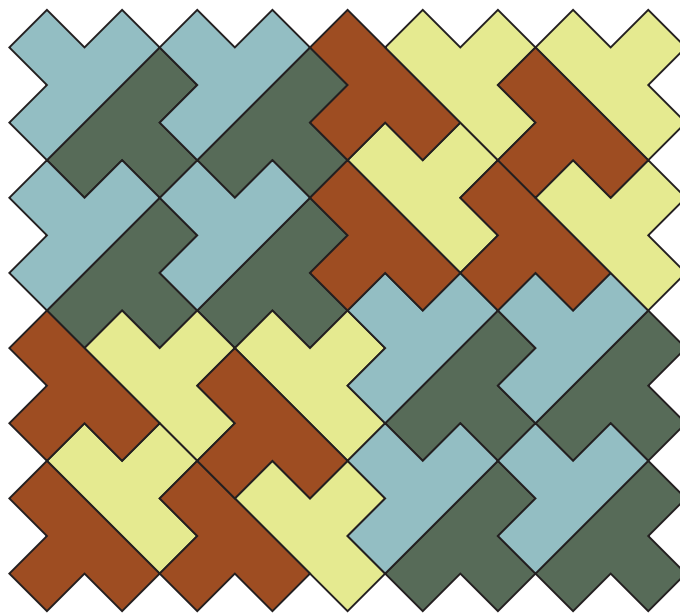


Figure 2.17 Tiling ragged rectangles with T-tetrominoes

3. Find examples of tilings of half-squares with tetrominoes that use as few as possible different types of tetrominoes.
4. Is it possible to tile a half-square with a combination of dominoes and straight trominoes?
5. What is the deficit of a half-square in a chessboard coloring?
6. What consequences does this deficit have? Can you argue that certain tilings are impossible?
7. Can you tile a half-square just with T-tetrominoes?

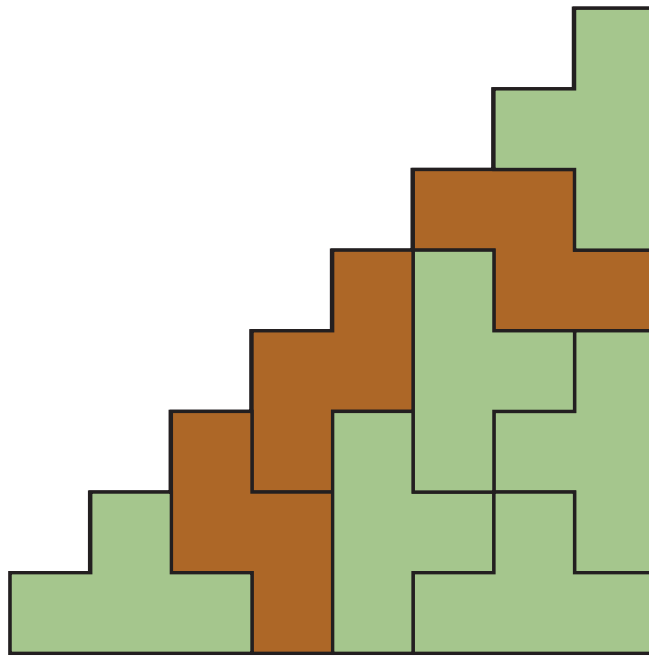


Figure 2.18 Tiling a ragged half-squares with S- and T-tetrominoes

2.2 POLYIAMONDS

A polyiamond is a polyform composed of equilateral triangles. The name is derived from the *diamond* shape, consisting of two equilateral triangles.

Name	Number of polyforms
Moniamond	1
Diamond	1
Triamond	1
Tetramond	3
Pentiamond	4
Hexiamond	12
Heptiamond	24
Octiamond	66

Table 2.2 none

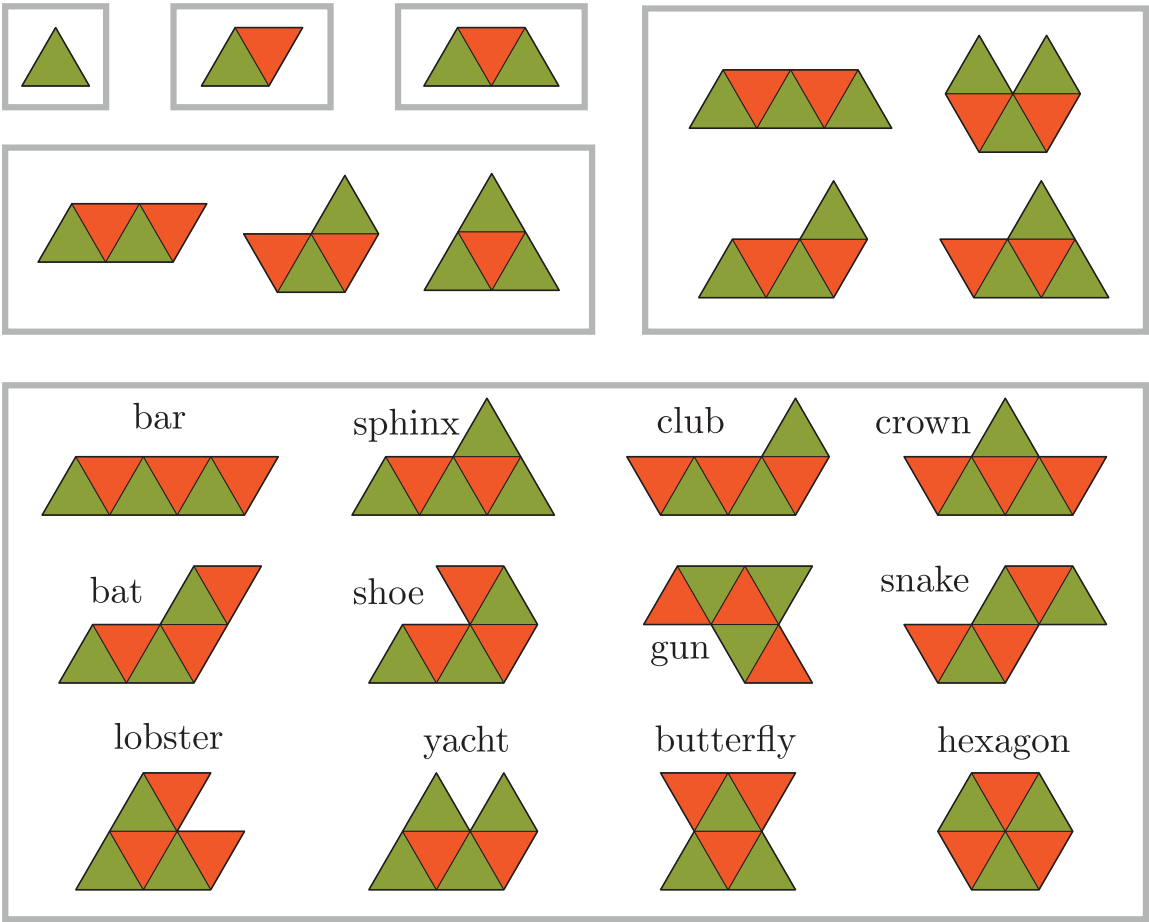


Figure 2.19 The monia-, dia-, tria-, tetria-, pentia- and hexiamonds

In the case of polyiamonds, we cannot expect to tile rectangles. But there are other attractive shapes that beg to be tiled: Triangles, trapezoids, rhombi, and hexagons, for instance. If we want to tile any such shape with congruent copies of a single tile, we first need to check whether its area is a multiple of that of the tile.

To this end, we will first gather some area formulas in a single proposition. All lengths are measured in multiples of the length of the base triangle, and all areas are measured in multiples of the area of the base triangle. Thus, computing the area of our shapes reduces to counting.

The only shape that is a little tricky is that of a hexagon with all angles being 120° . In **Figure 2.20** you can see an example. It also shows how to construct any such hexagon where the odd edges have given lengths and the opposite edges are longer by the same offset parameter d . We will see below that any 120° hexagon can be constructed this way.

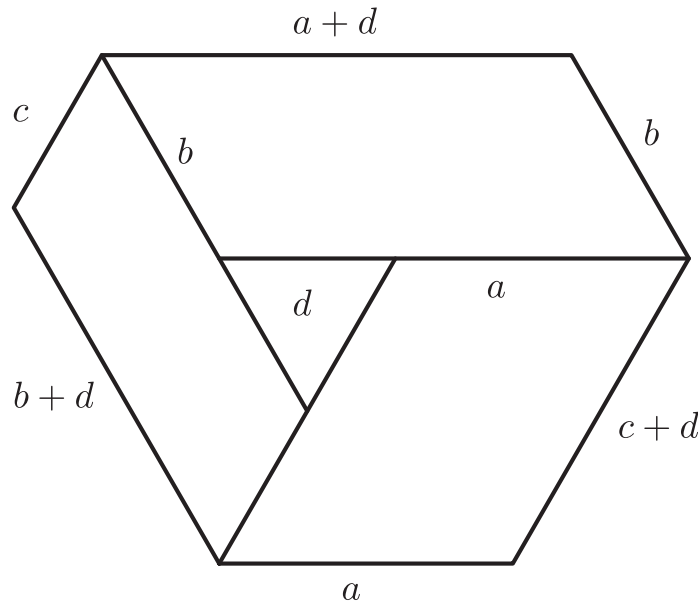


Figure 2.20 The general 120° hexagon

All polygons that can be tiled are limited in that their interior angles are multiples of 60° .

Proposition 2.2.1

1. For every $a > 0$, there is a unique equilateral triangle of side length a , and it has area a^2 .
2. For every $a > 0$ and $b > 0$, there is a unique parallelogram of side lengths a and b , and it has area $2ab$.
3. For every $a, b, c, d > 0$, there is a unique hexagon with three odd sides a, b, c and opposite even sides $a + d, b + d, c + d$. It has area $(a + b + c + d)^2 - a^2 - b^2 - c^2$.

Proof. The claim for the parallelogram is trivial, as any parallelogram can be skewed into a rectangle with the same side lengths. The claim for the triangle follows by dividing a parallelogram with $a = b$ into two triangles along the shorter diagonal.

For the hexagon, we begin by showing that any 120° hexagon is given as in the figure. To see this, take an arbitrary 120° hexagon, and label the edges counterclockwise from 1 to 6. Choose the labeling in such a way that the total length of the odd edges is not greater than the total length of the even edges. We can, by modifying our choices if necessary, also assume that the edge 2 is at least as long as the opposite edge 5. Denote the lengths of the off edges 1, 3, 5 by a, b, c , respectively. Finally, let $c + d$ be the length of edge 2, with $d \geq 0$. We have to show that edges 4 and 6 have lengths $c + d$ and $b + d$, respectively.

Now complete the hexagon to a triangle by adding equilateral triangles to the edges 1,3,5. The result is a 60° triangle whose edge that extends the edge 2 has length $a + b + c + d$. But all 60° triangles are equilateral, so in fact all sides of the triangle have the same length. This determines the lengths of edges 4 and 6 as claimed.

We obtain the formula for the area as $(a + b + c + d)^2 - a^2 - b^2 - c^2$ by subtracting the areas of the three attached triangles from the area of the large triangle. $\square$

Let's start tiling these shapes with diamonds. Can we tile a triangle of side length a with a diamond? Clearly, the area must be even for that, so a should be even as well. But that doesn't help — no triangle can be tiled with diamonds. We give two arguments for that.

Proposition 2.2.2 *No shape that has an edge with two acute corners can be tiled with diamonds.*

Proof.

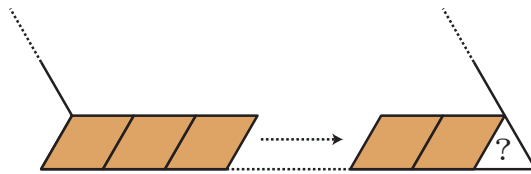


Figure 2.21 Domino effect with a diamond

An acute corner forces the placement of a diamond with one triangle in that corner. What remains of the shape to be tiled still has two acute corners along the same edge. Proceeding along that edge, we end up with a single untiled triangle at the other acute vertex, as shown in **Figure 2.21**. $\square$

This condition is somewhat local in nature — we do not care how the shape looks away from the edge with the two acute corners. Our second argument uses the coloring method. We can 2-color the triangle tiling by coloring all 'up'-triangles in one color, and all 'down'-triangles with the other color. Then a diamond is balanced, as it covers one triangle of each kind.

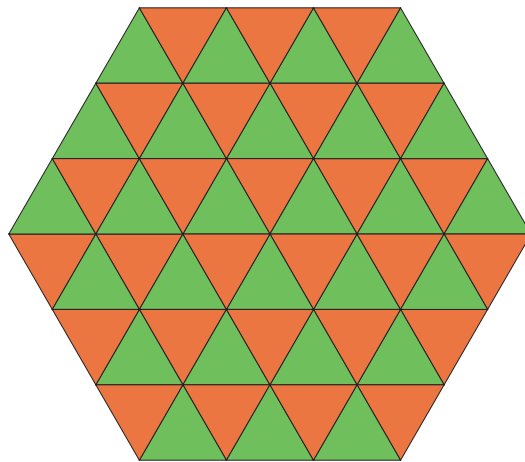


Figure 2.22 2-coloring of the triangle tiling

Now suppose we have an ‘up’-triangle with edge length a . The number of ‘up’-triangles within it is the triangle number $\frac{1}{2}a(a+1)$, and the number of ‘down’-triangles the next smaller triangle number $\frac{1}{2}a(a-1)$. Thus, the *unbalance* of our big triangle is the difference

$$\frac{1}{2}a(a+1) - \frac{1}{2}a(a-1) = a$$

In particular, all triangles are unbalanced, and we have our second reason why triangles cannot be tiled by diamonds.

On the other hand, any parallelogram can be obviously tiled with diamonds, so we next turn our attention to hexagons. These have only obtuse corners, which is a good sign. Let's determine the unbalance of a hexagon as in **Figure 2.20**, with side lengths a, b, c , and excess d .

This figure shows the hexagon decomposed into three parallelograms and a ‘down’-triangle of side length d . The parallelograms are balanced, so the total unbalance of our hexagon is that of the triangle, namely $-d$.

This proves:

Proposition 2.2.3 *A hexagon is balanced if and only if $d = 0$.*

Even better: We also see from this decomposition that any hexagon with $d = 0$ can be tiled with diamonds, as the three parallelograms surely can.

What does this 2-coloring tell us about other polyiamonds? The triamond is unbalanced and happily tiles a triangle, rhombus, and hexagon of side length 3 each.

The first two tetriamonds are unbalanced, while the triangle tetriamond tiles a triangle, rhombus, and hexagon of side length 4 each.

Exploration 2.2.1 We will explore what shapes can be tiled with the four pentiamonds. We also introduce a new shape to tile, the *star*: This is a regular hexagon of edge length a with a triangle of edge length $b \leq a$ attached to each side so that the whole shape is rotational symmetric, as in **figure 2.23**.

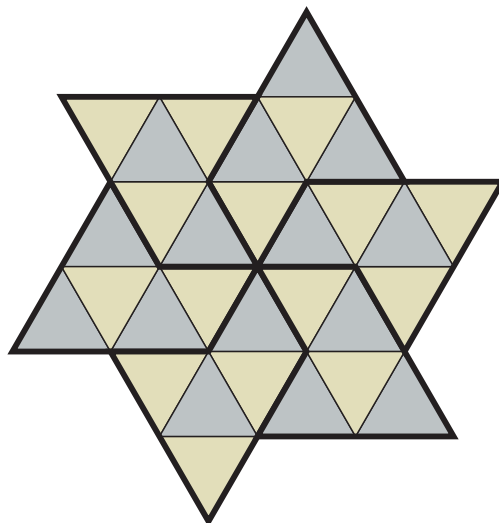


Figure 2.23 Pentiamond tiling of a star

1. Which parallelograms can be tiled with the I-tetriaomond?
2. Can a triangle be tiled with the I-tetriaomond?
3. Which hexagons can be tiled with the I-tetriaomond?

4. Determine the area of a star, depending on the sizes of the base hexagon and triangles.
5. Find tilings of stars by one or several pentiamonds.
6. Find tilings of hexagons by one or several pentiamonds.

The hexiamonds have been named, as indicated in **Figure 2.19**. All but the yacht and the sphinx of the hexiamonds are balanced and hence will not tile any triangle.

For the yacht, the argument in **Proposition 2.2.2** works as well and also shows that this hexiamond won't tile any triangle. The remaining candidate is sphinx, which indeed can tile a triangle. The smallest solution, due to BRENDAN OWEN, tiles a triangle of side length 12 and is shown in **Figure 2.24**.

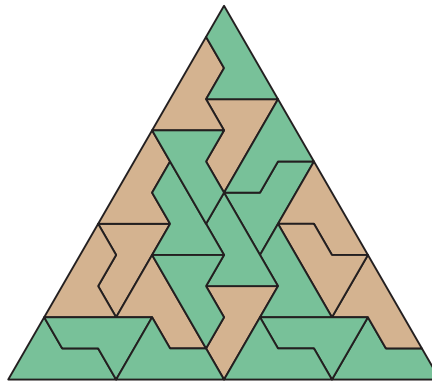


Figure 2.24 Hexiamond Tiling of a Triangle

Exercise 2.2.4 Use the 2-coloring to show that the sphinx cannot tile a triangle of size 6.

Solution: Suppose we had a tiling of the size 6 triangle by 6 sphinxes. The unbalance of the triangle is 6, and each sphinx has an unbalance of +2 or -2. Thus, the total unbalance of the sphinxes is divisible by 4, which is a contradiction. This argument works for all triangles of size $6n + 6$.

There are no known general procedures to find tilings like these, but more or less sophisticated variations of brute force. On the other hand, because we know that the sphinx tiles a triangle, we have no good reason to look for a coloring argument that would disprove this.

More flexibility is possible if we allow more than one polyiamond in a tiling. Still, the coloring argument gives us hints about what a tiling needs to look like.

For instance, when tiling an 'up'-triangle of side length $2a$ with any combination of a^2 tetriamonds, we will need a more of the 'up'-triangle-tetriamonds than of the 'down'-triangle-tetriamonds.

Exercise 2.2.5 Find three different tilings of the size 6 triangle, using any combination of tetriamonds.

Next, we will look at a coloring with three colors. This one is most suited for polyiamonds consisting of a number of triangles divisible by 3.

Exercise 2.2.6 Determine which of the hexiamonds are 3-balanced, i.e., cover two triangles of each color, no matter how they are placed.

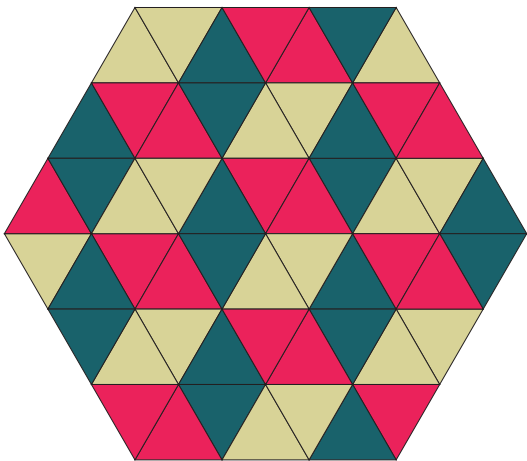


Figure 2.25 Three-coloring of a triangle tiling

<i>name</i>		<i>colors covered</i>
bar		2 + 2 + 2
club		1 + 2 + 3
crown		2 + 2 + 2
sphinx		1 + 2 + 3
snake		2 + 4 or 1 + 2 + 3
bat	1 + 1 + 4 or 2 + 2 + 2 or 3 + 3	
gun		1 + 2 + 3
lobster		2 + 2 + 2
hook		1 + 2 + 3
hexagon		2 + 2 + 2
butterfly		2 + 2 + 2

Table 2.3 none

Solution: We give a list of possible color distributions. For instance, 1 + 2 + 3 means that the tile covers one color once, another color twice, and the third color three times.

Consequently, the bar, crown, lobster, hexagon and butterfly all all 3-balanced with respect to this tiling.

Exercise 2.2.7 Show that any shape that is symmetric with respect to a 120° rotation about a vertex is 3-balanced.

Exercise 2.2.8 Eight of the twelve hexiamonds can tile a regular hexagon of some size. Find them, and give an example in each case.

Very little is known positively about tilings by polyiamonds that goes beyond searching all possibilities. The following construction is due to ANDREW CLARKE:

Definition 2.2.9 An (r, s) -hook is a polyiamond consisting of $2r + 2s$ triangles such that the sets of $2r$ and $2s$ triangles form a rhombus of short side length 1 and of long side lengths r and s , respectively, which are joined at a 120° angle.

For instance, the hook hexiamond is a $(1, 2)$ -hook.

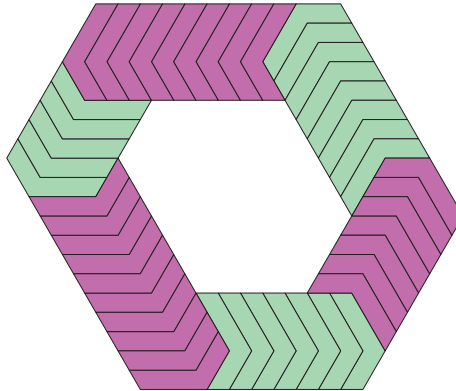


Figure 2.26 Tiling hexagonal annuli with (r, s) -hooks

Proposition 2.2.10 An (r, s) -hook can tile any hexagonal annulus of width $r + s$.

Proof. Form a straight chain, and bend and flip at the right places, as shown in **Figure 2.26**. □

Corollary 2.2.11 If a (r, s) -hook tiles an (a, b, c, d) -hexagon, it also tiles an $(a + r + s, b + r + s, c + r + s, d)$ -hexagon.

Example 2.2.2 The $(1, 2)$ -hook tiles any regular hexagon except for side lengths 1 and 4. To see this, note that the hook tiles easily a regular hexagon of side lengths 2 or 3, and not so easily a regular hexagon of side length 7.

Exercise 2.2.12 Find examples of small (not necessarily regular) hexagons that can be tiled by hooks. This exercise is clearly open-ended.

Exercise 2.2.13 Show that the hexagon in **Figure 2.28** can be tiled with diamonds if any of the yellow triangles is removed.

Solution:

Any two consecutive triangles along the path in **Figure 2.29** can be covered by a diamond. If we remove one yellow triangle from the hexagon, the path breaks apart into pieces of even length, each of which can be tiled with diamonds.

Exploration 2.2.3 A snowflake Σ_n of order n is constructed as follows. Σ_1 is just a triangle of side length 1. Σ_2 is a triangle of side length 3 to whose sides we attach triangles with side length 1, yielding a hexagram or six-pointed star. Σ_3 starts with a triangle of side length $3^2 = 9$, to which side are attached triangles of side length 3, with finally attached triangles of size 1 to the middle of all segments of length 3. Higher-order snowflakes are constructed using the same recipe inductively.

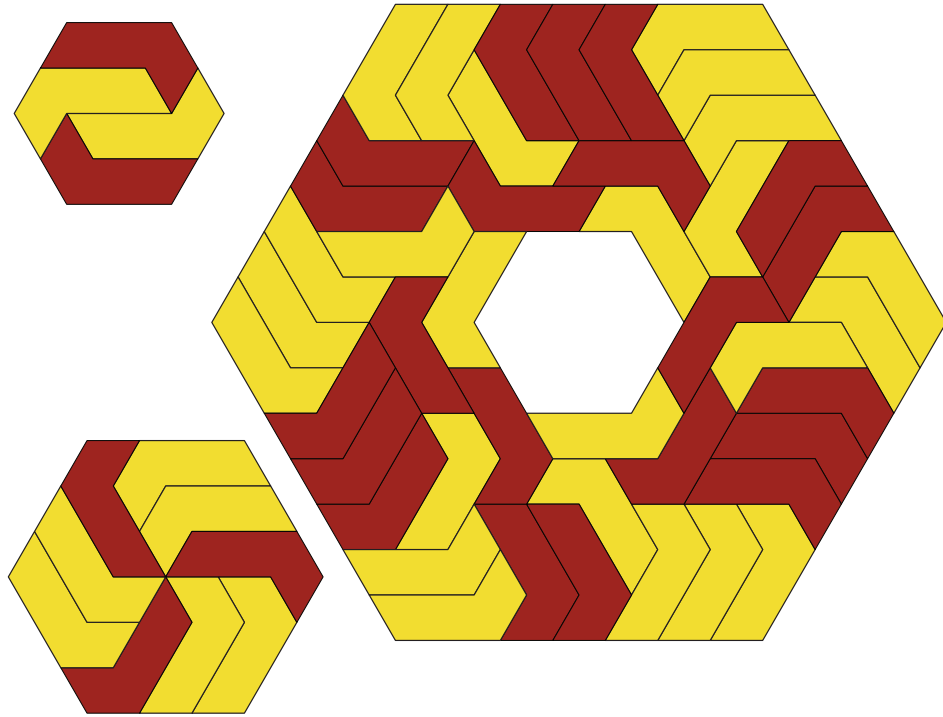


Figure 2.27 Tiling hexagons with hooks

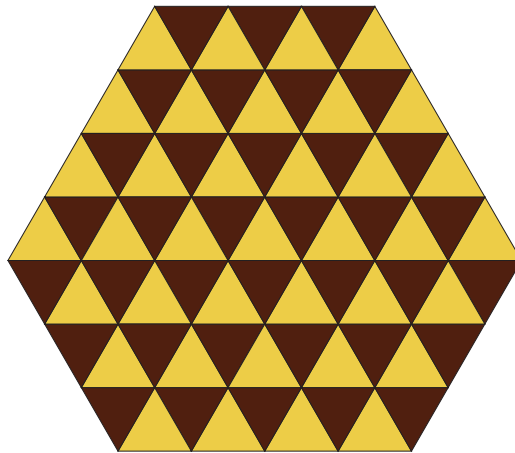


Figure 2.28 Tiling deficit
hexagons with diamonds

1. Compute the area of Σ_n in multiples of a unit triangle.
2. Count how many triangles of each color the snowflake Σ_n covers in the 2-coloring and 3-coloring, respectively. Does this depend on how you place the snowflake over the periodic colored triangle tilings?
3. Find polyiamond tilings of Σ_n with small polyiamonds.
4. Can you show that certain small polyiamonds do not tile a snowflake Σ_n ?

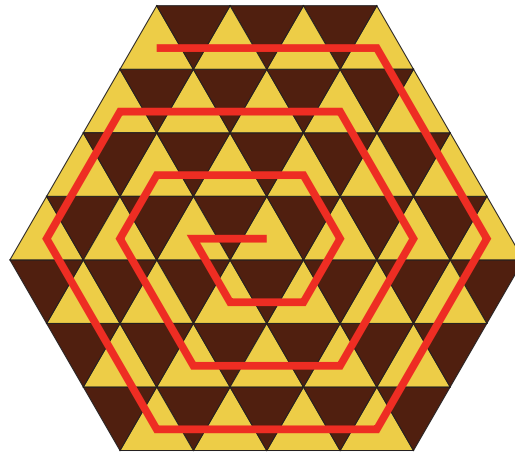


Figure 2.29 Tiling deficit hexagons with diamonds

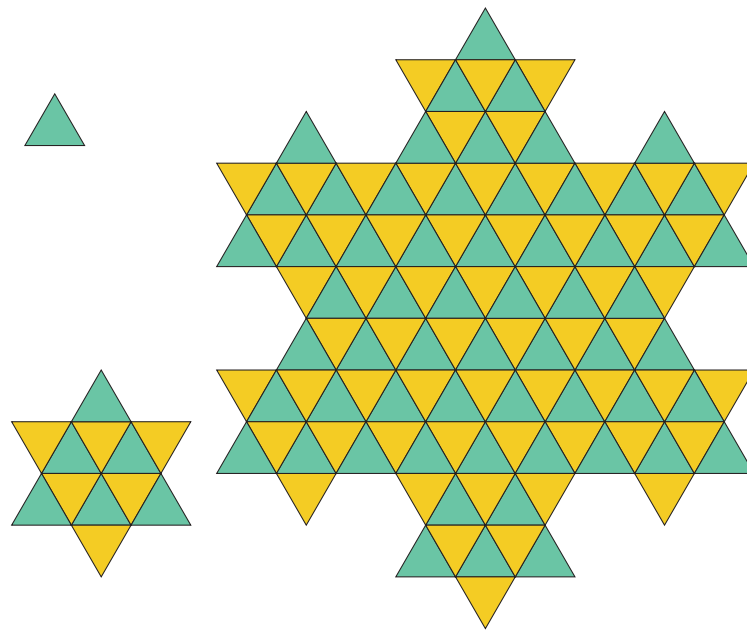


Figure 2.30 Snowflakes of order 1, 2, and 3

5. It is easy to tile Σ_3 with the arrow tetriamond. However, the arrows don't like to line up, i.e. to share an edge and face in the same direction. Find a tiling of Σ_3 where this doesn't happen.
6. Find a tiling of Σ_3 by arrows where as many arrows as possible point in the same direction.

Solution: See Figure 2.31.

Exploration 2.2.4 We are going to tile ragged hexagons with a selection of hexiamonds. In a ragged hexagon, the sides zigzag as in Figure 2.32, which shows a ragged hexagon of dimensions 3, 4, 3, 4, 3, 4. Note that the boundary changes directions at every unit step, and makes only 120° angles. You can also think of a ragged hexagon as one of the hexagons we will discuss in Section 2.3.

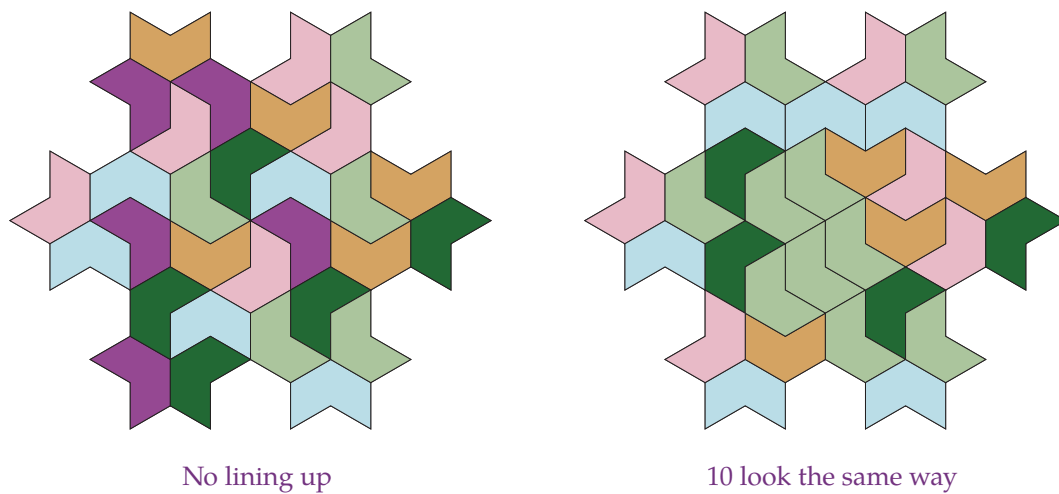


Figure 2.31

The hexiamonds we are going to use arise from starting with the hexagon hexiamond, removing one of its triangles, and adding it elsewhere to the hexiamond. Up to symmetry, there are three of these, which I will call Sharp, Soft, and Straight, as indicated in the figure. Straight has its removed triangle attached to the opposite edge, Sharp to the adjacent edge, and Soft to the edge in between the adjacent and the opposite edge. Please conduct an investigation, trying to answer the following questions.

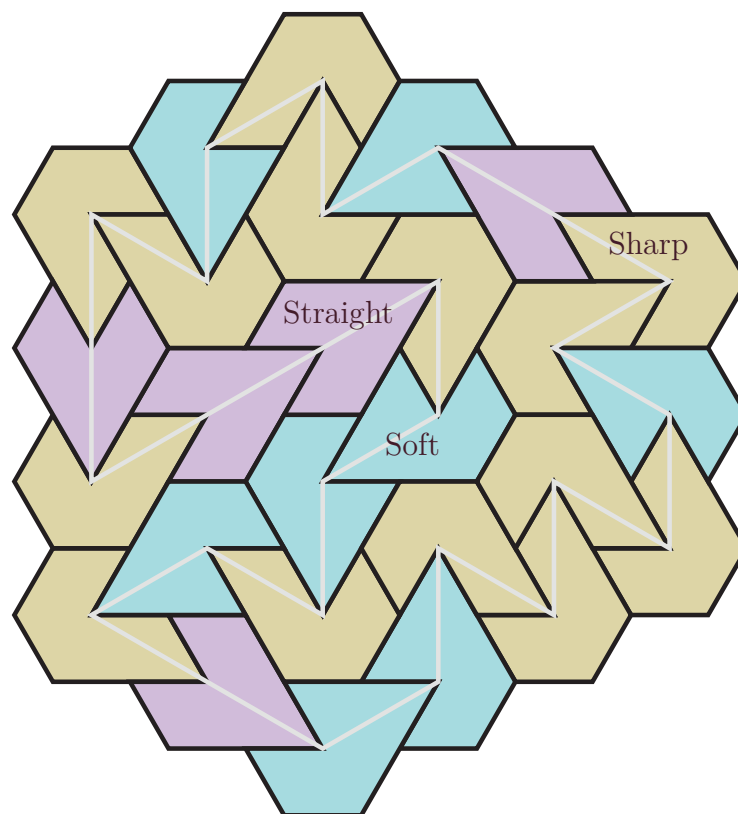


Figure 2.32 A ragged hexagon, tiled with Sharp, Soft, and Straight.

1. What is the area of a ragged hexagon of edge lengths $a, c + d, b, a + d, c, b + d$? Find a formula and test it on

examples. Explain why you think the formula is correct. It might help to look at [Lemma 2.3.1](#)

2. Find examples of tilings of ragged hexagons with Sharp, Soft, and Straight different from the one above.
3. Can one tile a ragged hexagon with Sharp alone? With Soft alone or with Straight alone? Give arguments!
4. What do the coloring techniques we discussed for polyiamonds tell us here?
5. Can you tile a ragged hexagon with Sharp and Straight? With Soft and Straight?
6. There is a curious grayish path drawn in [Figure 2.32](#) that connects the ‘tips’ of each of Sharp, Soft, and Straight with the tip of the next piece into which it ‘fits’. Can you use that?
7. Find a ragged hexagon that can be tiled with Sharp and Soft alone.
8. Find a ragged hexagon that can be tiled with Sharp.

Solution: See [Figure 2.33](#).

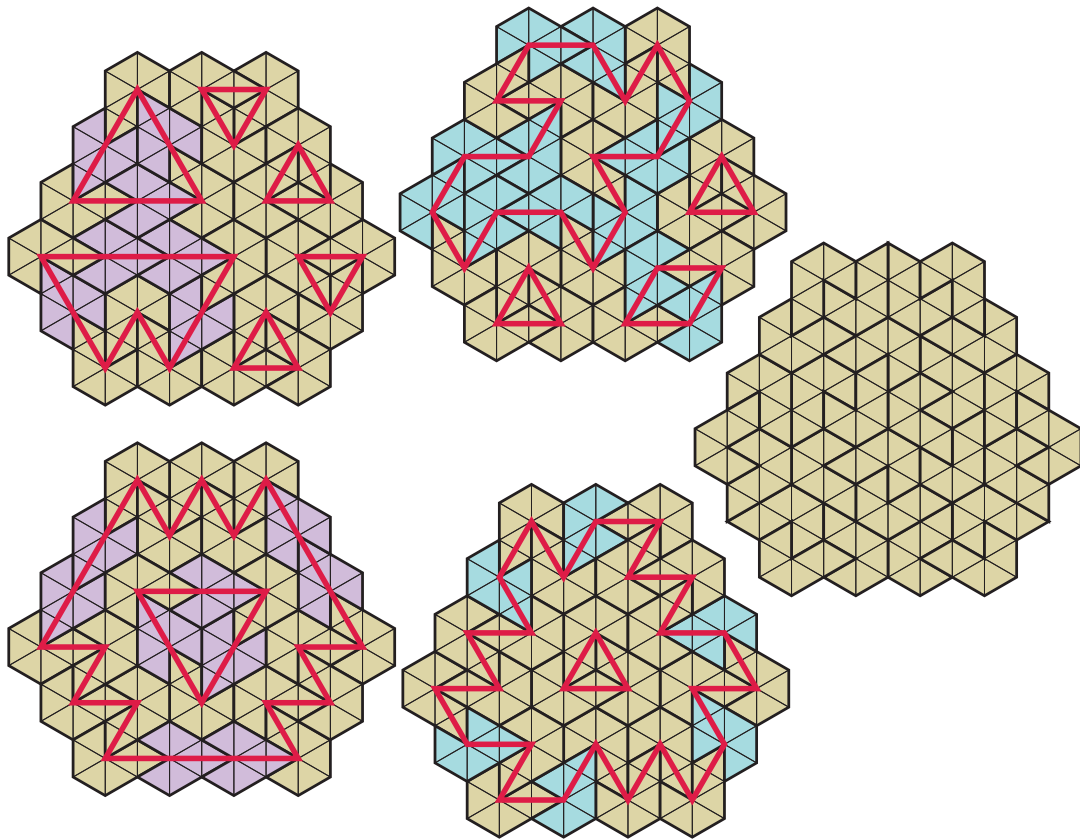


Figure 2.33 The solution just with Sharp was found by ALAN SCHOEN

2.3 POLYHEXES

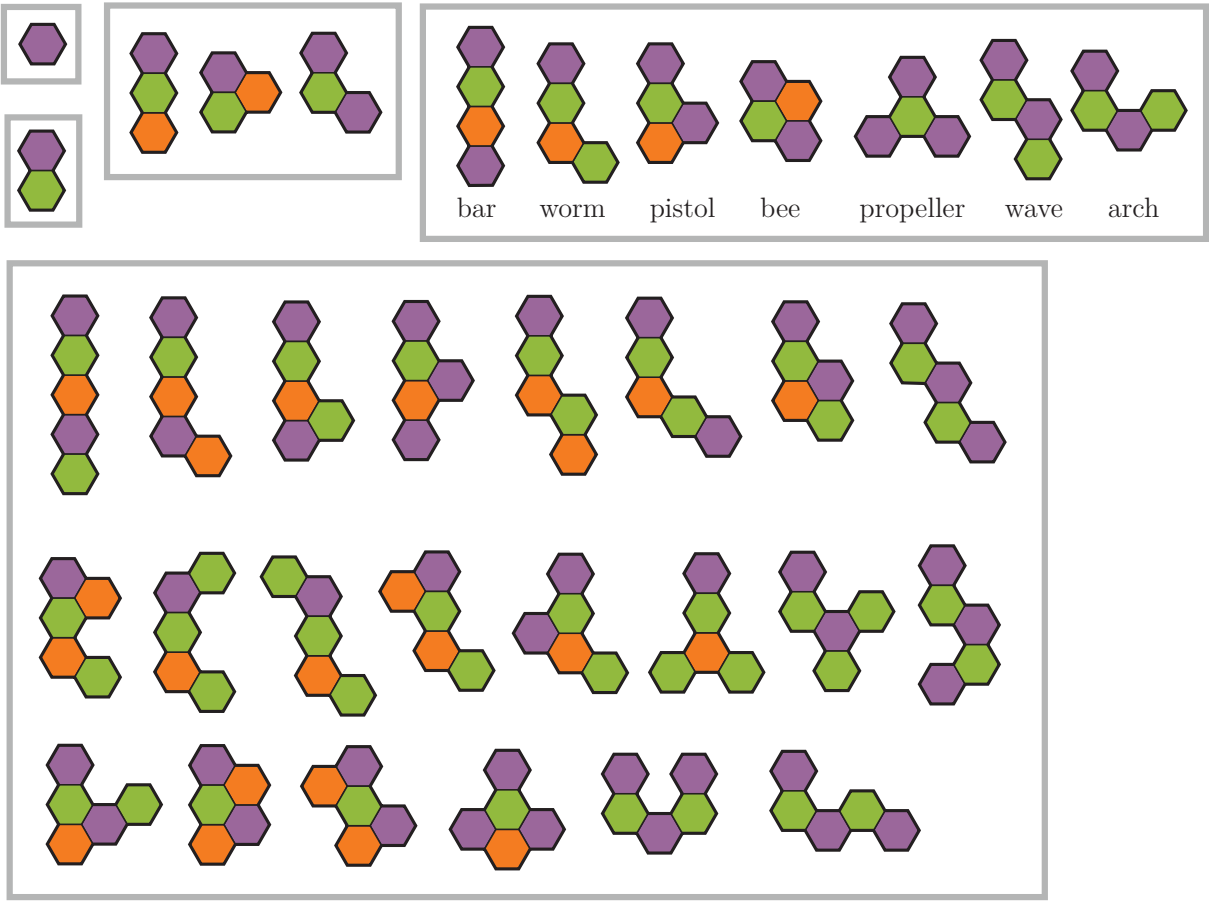


Figure 2.34 The mono-, di-, tri-, tetra-, and pentahexes

The only regular polygons that tile the plane are equilateral triangles, squares, and regular hexagons. We have discussed the first two cases, which lead to polyominoes and polyiamonds. The polyforms that arise from gluing regular hexagons together are called *polyhexes*.

Here are their names and numbers:

<i>name</i>	<i>number</i>
Monohex	1
Dihex	1
Trihex	3
Tetrahex	7
Pentahex	22
Hexahex	82

Table 2.4 none

We first compute the area of some basic shapes. We measure length and area in multiples of the width and area of our unit hexagon. Note that the formulas are different from the ones we found for polyiamonds.

Proposition 2.3.1

1. A triangle with side length a has area $\frac{1}{2}a(a+1)$.
2. A hexagon with side lengths a, b, c and excess d has area $\frac{1}{2}(a+b+c+d-2)(a+b+c+d-1) - \frac{1}{2}a(a-1) - \frac{1}{2}b(b-1) - \frac{1}{2}c(c-1)$.
3. A regular hexagon of side length a has area $3a(a-1)+1$.

Proof. The triangle area is just a triangle number. For the hexagon area, we use the same method as in the section about polyiamonds: We extend the hexagon to a large triangle by adding small triangles to the short edges. Note that the triangle we add to a side of length a will have side length $a-1$ only. Therefore, our long triangle will have side length $a+b+c+d-2$, and the sides of the added triangles will be $a-1$, $b-1$, and $c-1$. The area of the hexagon is that of the large triangle, and less than the area of the extension triangles. $\square$

As we can see in **Figure 2.35**, triangles admit a Hamiltonian circuit. In particular, if the triangle has an even number of hexagons, it can be tiled with dihexes.

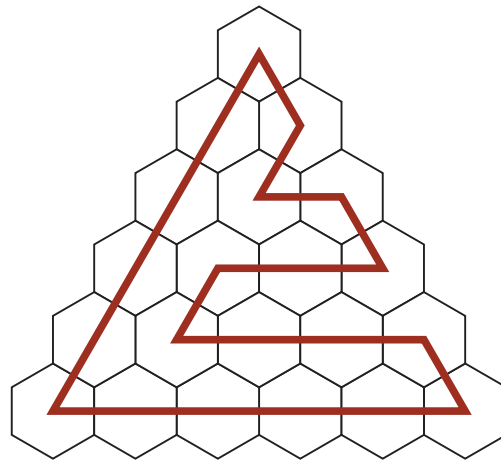


Figure 2.35 Hamiltonian circuit in a triangle

Exercise 2.3.2 Take a triangle of side length 10. Show that you can remove an arbitrary hexagon and tile the defective triangle with 27 dihexes.

Solution: Draw a Hamiltonian circuit in the triangle as in **Figure 2.35**. Removing a single hexagon leaves an open Hamiltonian circuit of length 54, which can be covered with dihexes from end to end.

There are many results, both positive and negative, about the tilability of triangles and hexagons by special polyhexes. Many positive results use a divide and conquer method: One first finds a collection of geometrically simple shapes that the polyhex can tile, and then uses these to tile larger shapes. To explain this, we first introduce a bit of notation: We denote by T_n a triangle of side length n . Thus T_1 is the monohex, T_2 the triangular trihex, and T_3 the triangular hexahex. By $R(a, b)$ we denote a trapezoid with the shorter of the two parallel sides being of length a and the length of the non-parallel sides being of length b .

We will use the following decompositions:

Lemma 2.3.3

1. If we can tile $T(a)$ and $R(a+1, b)$, then we can also tile $T(a+b)$.
2. If we can tile $T(a-b)$ and $R(a+1, b)$, then we can also tile $T(a+2b)$.

Proof.

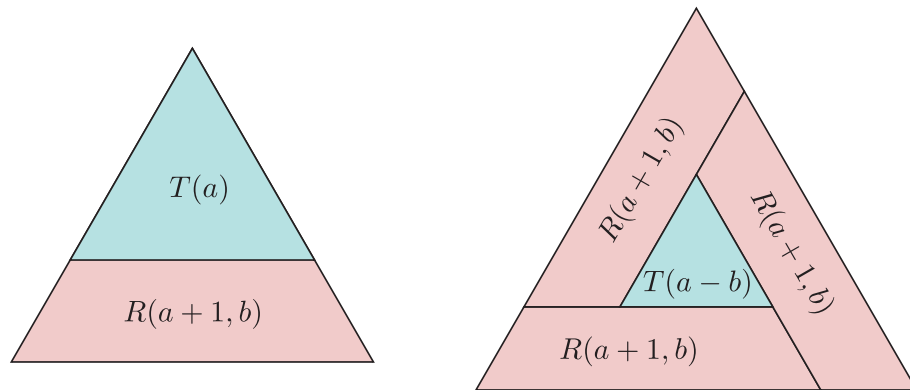


Figure 2.36 Divide and conquer

The figure illustrates how we decompose the larger shape into smaller shapes. The only difficulty is that adjacent edges do not necessarily have the same length in hexagon units. $\square$

We illustrate this method with the triangular trihex:

Lemma 2.3.4 *The triangular trihex $T(2)$ tiles the trapezoids $R(2k, 3)$ and $R(3k+1, 2)$ for any k .*

Proof. See **Figure 2.37**.



Figure 2.37 Tiling thin rhombi with $T(2)$

Corollary 2.3.5 *The triangular trihex $T(2)$ tiles all triangles $T(n)$ with n of the form $12k$, $12k+2$, $12k+9$, and $12k+11$.*

Proof. Using rhombi of the form $R(3k+1, 2)$, the two previous lemmas allow us to extend a tiling of $T(3k)$ to a tiling of $T(3k+2)$. Secondly, with rhombi of the form $R(2k, 3)$ we can extend a tiling of $T(2k-1)$ to a tiling of $T(2k+2)$, and a tiling of $T(2k)$ to a tiling of $T(2k+9)$.

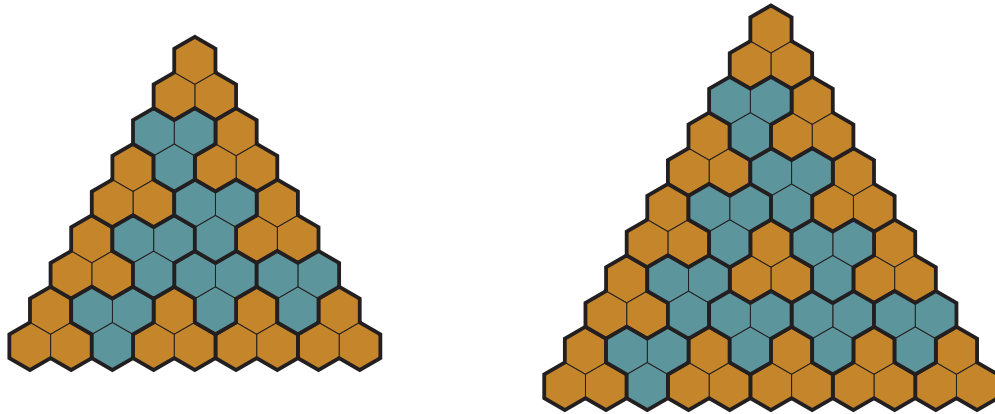


Figure 2.38 Tiling Triangles of size 9 and 11 with $T(2)$

Thus, given a tiling of the form $T(12k)$, we can extend it in one step to tilings of $T(12k + 2)$ and $T(12k + 9)$. In a second step, these can be extended to tilings of $T(12k + 11)$ and of $T(12(k + 1))$. Starting with $T(0)$, this allows us to tile all triangles inductively as claimed. $\square$

JOHN CONWAY and JEFF LAGARIAS have shown that these are indeed the only triangles that can be tiled by $T(2)$, see [6,7].

Exercise 2.3.6 Find geometrically different tilings of $T(9)$ and $T(11)$ by $T(2)$ than the ones shown in **Figure 2.38**.

Exercise 2.3.7 Show that the pistol tetrahex can tile any $T(8k)$ and $T(8k - 1)$, but no other $T(n)$.

Tiling regular hexagons of side length n with small polyhexes is impossible, as the area $3n^2 - 3n + 1$ is never divisible by 2, 3, or 5. Tilings with heptahexes have been well studied, however, mostly with the help of computers.

To show that certain shapes cannot be tiled, coloring arguments are again a very effective tool. We will discuss two simple colorings of the hexagon tiling.

Definition 2.3.8 A shape made up of hexagons is 3-balanced if it contains the same number of hexagons of each color in the 3-coloring of the plane in **Figure 2.39**, no matter how the shape is placed.

As an immediate consequence, we have:

Lemma 2.3.9 The I- and D-trihex can only tile balanced shapes.

Exercise 2.3.10 Find tilings of $T(3k)$ and $T(3k - 1)$ with both I- and D-trihexes.

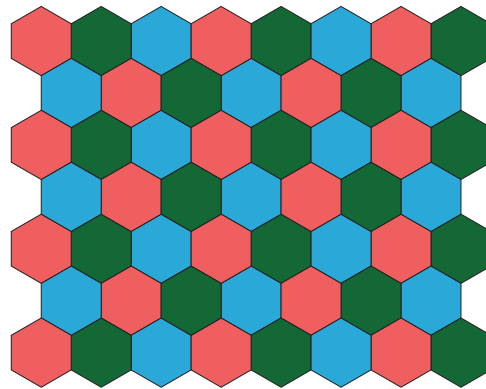


Figure 2.39 3-coloring of the hexagon tiling

Exercise 2.3.11 Is a 120° hexagon whose area is divisible by 3 automatically 3-balanced?

Exercise 2.3.12 Find all hexahexes that cover just three hexagons of two different colors of the 3-coloring, and formulate a theorem giving a condition for tiling a shape with these hexahexes.

Sometimes, coloring arguments can be used even if the tiles are not balanced. Here is an example:

Theorem 2.3.13 The C- and S-tetrahex cover an even number of hexagons each color, and the same must hold for any region they tile.

Our second coloring of the hexagon tiling uses four colors.

Definition 2.3.14 A shape made up of hexagons is 4-balanced if it contains the same number of hexagons of each color in the 4-coloring of the plane in **Figure 2.40**, no matter how the shape is placed.

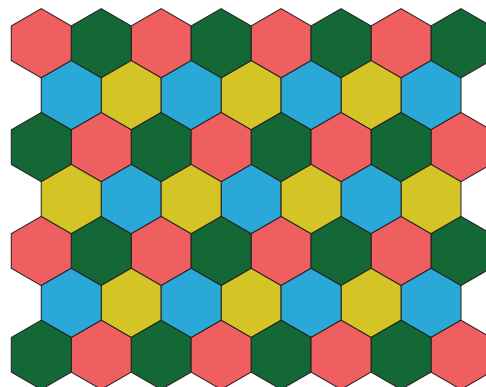


Figure 2.40 4-coloring of the hexagonal tiling

Theorem 2.3.15 *The O-, Y-, and S-tetrahexes are 4-balanced, and so is any region they tile.*

Exercise 2.3.16 *Find all hexahexes that cover just three hexagons of two different colors of the 4-coloring, and formulate a theorem giving a condition for tiling a shape with these hexahexes.*

Exploration 2.3.1 We will explore polyforms that are made up of squares and 45° rhombi. A *polyrhomb* of type (r, s) consists of r such rhombi and s squares, glued together edge-to-edge, without overlaps. In particular, a polyrhomb of type $(0, n)$ is just an n -omino.

1. Enumerate all dirhombs, i.e. polyrhombs of type (r, s) with $r + s = 2$.
2. Enumerate all trirhombs, i.e. polyrhombs of type (r, s) with $r + s = 3$.
3. Can you tile a regular octagon of edge length 1 and 2 with trirhombs? Try to use as few different types of trirhombs as possible.
4. Determine for all your tilings the types (r_i, s_i) of the polyrhombs you used. Compute $\sum r_i$ and $\sum s_i$. Do you recognize a pattern? Can you find a reason for the pattern?
5. Can you tile a square of edge length n with polyrhombs that are not just squares? How many different types do you need?

3 Dissection Puzzles and Periodic Tilings

In this chapter, we study tilings of the entire Euclidean plane that repeat themselves. We begin by introducing the proper language from Linear Algebra to make periodicity and other symmetries precise and to compute with them. Then we will classify the simplest periodic tilings and discuss a useful criterion for tiling periodically.

Next, we will use our knowledge of periodic tilings to tackle a seemingly unrelated problem: That of finding effective ways to dissect one shape into another. We will end this chapter by looking at theorems from Geometry that have simple explanations using tilings and dissections.

3.1 PARADOXICAL DISSECTIONS

In this introductory section, we will discuss dissections that pretend to change the area of a shape by cleverly disguising the flaws.

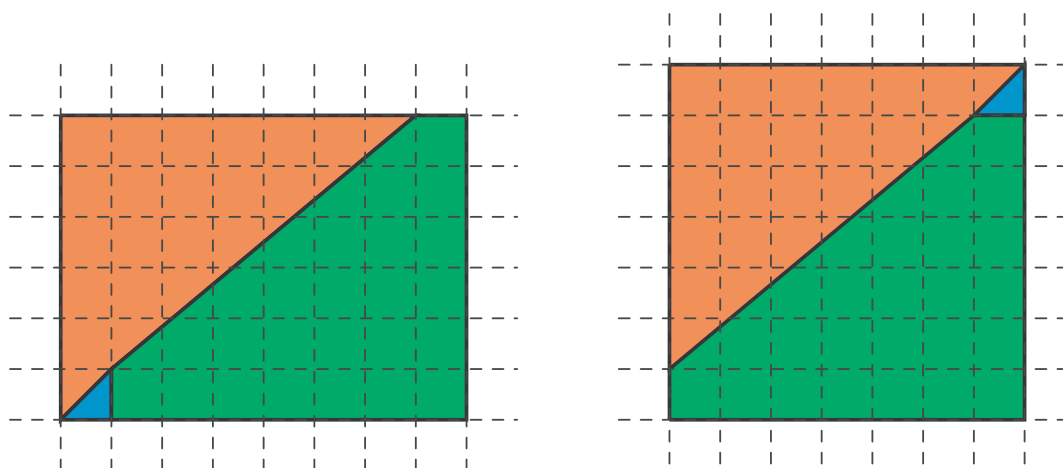


Figure 3.1 Creating Matter?

Exercise 3.1.1 A 6×8 rectangle is dissected into three pieces as shown in Figure 3.1, and reassembled into a 7×7 square, increasing the area by one unit square. How can this be?

There are many variations of this trick. The following one is a little more subtle:

Figure 3.2 seems to show that one can dissect a 13×13 square into an 8×21 rectangle, thus diminishing the area by one unit square. This shouldn't be possible, but where is the mistake?

The flaw becomes more obvious when we look at the same dissection with fewer pieces. In figure 3.3 we decompose an 8×8 square into a 5×13 rectangle, increasing the area by one unit square.

But here the long diagonal of the rectangle at the right doesn't look quite right anymore, and in fact it isn't: The slopes of the three segments are $\frac{2}{5}$, $\frac{1}{3}$ and $\frac{2}{5}$, so the diagonal is not straight.

The same explanation works for the first example, only here the slopes are $\frac{3}{8}$ and $\frac{2}{5}$.

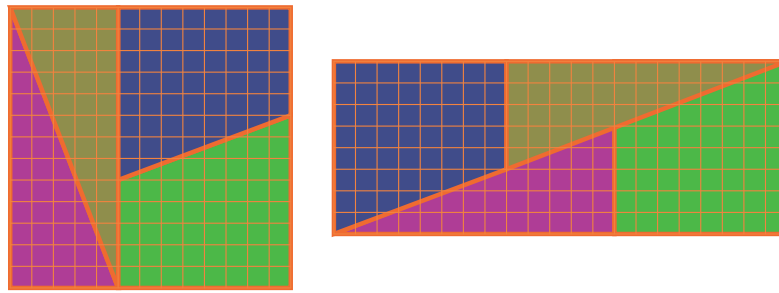


Figure 3.2 Fibonacci does it again

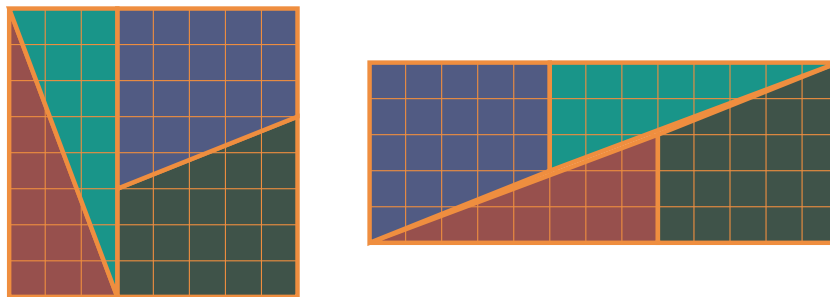


Figure 3.3 A broken diagonal

The reason why these paradoxical decompositions work so well with the chosen dimensions of the rectangles lies a bit deeper. The edge lengths of the tiles are elements of the sequence

$$2, 3, 5, 8, 13, 21, \dots$$

which are the famous *Fibonacci numbers*, defined by

$$F_1 = F_2 = 1$$

$$F_{n+2} = F_n + F_{n+1}$$

These numbers satisfy the formula

$$F_n^2 = F_{n-1} \cdot F_{n+1} \pm 1.$$

This explains why the areas of a rectangle and a square differ by only one — the least suspicious amount.

Also, the ratio F_n/F_{n+1} of consecutive Fibonacci numbers converges rapidly to $1/\phi \approx 0.618034$, the reciprocal of the golden ratio. This explains why the slopes of the segments of the diagonal are almost the same.

golden
ratio

Exercise 3.1.2 The following figure shows a dissection of a triangle into six tiles, which are then rearranged into figures with a smaller area. Note that all tiles have corners on the intersection of the grid lines. Identify the congruent pieces by placing numbers from 1 to 6 into the tiles in each figure. Compute the area of the region that the tiles are suggested to cover in each case. Explain the flaw by computing and indicating the slope of all relevant lines.

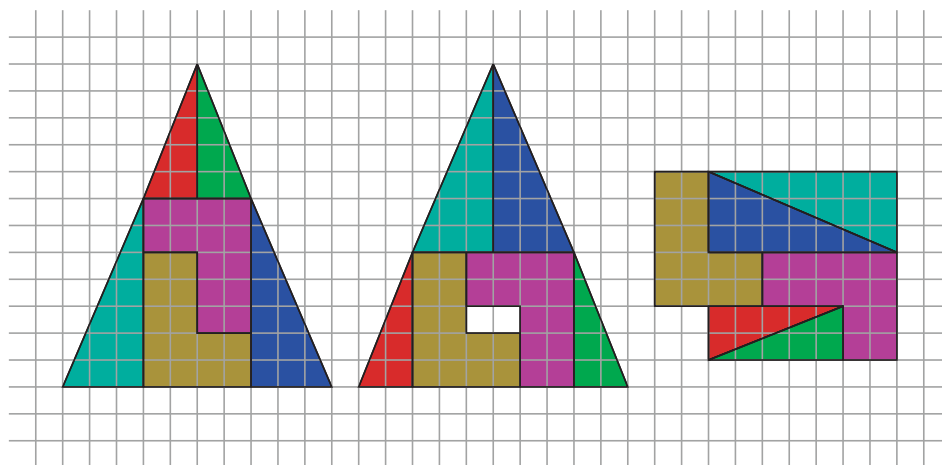


Figure 3.4 More dissection trouble

3.2 A TASTE OF LINEAR ALGEBRA

Often, our tilings will have symmetries, and the simplest symmetries are *translations*.

In coordinates, a translation is expressed as a *vector* $v = \begin{pmatrix} x \\ y \end{pmatrix}$ that tells by what amount the x - and y -coordinates of the points of a shape change when the translation is applied. Even though translations just look like points of the plane in coordinates, they are, strictly speaking, *differences* between points.

We can add two translations $v_1 = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ by adding the coordinates:

$$v_1 + v_2 = \begin{pmatrix} x_1 + x_2 \\ y_1 + y_2 \end{pmatrix}.$$

Geometrically, this means that we first translate using v_1 and then using v_2 .

We can also multiply a translation $v = \begin{pmatrix} x \\ y \end{pmatrix}$ by any real number c , again just by multiplying both coordinates:

$$cv = \begin{pmatrix} cx \\ cy \end{pmatrix}.$$

Geometrically, we represent translations by arrows and note that parallel arrows represent the same translation, as shown in **Figure 3.5**.

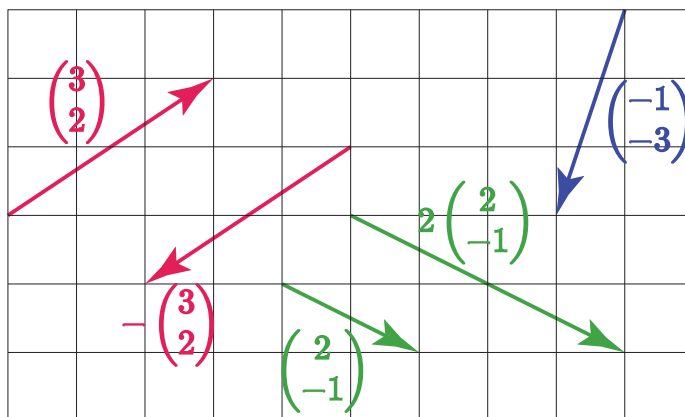


Figure 3.5 Coordinate notation for vectors

A translation can not only be used to shift individual points, but entire sets (i.e., shapes). We define:

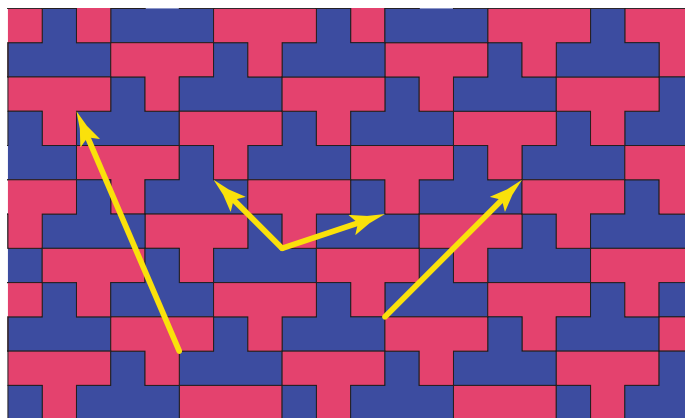
Definition 3.2.1 The translation of a shape S by a vector v is the set $S + v = \{p + v : p \in S\}$. A shape S is invariant under a translation v if $S = S + v$. A tiling is invariant under a translation if the translation shifts each tile to another tile of the same tiling.

Example 3.2.1 The set $S = \{(a, b) \in \mathbb{R}^2 : a, b \in \mathbb{Z}\}$ is invariant under the translations $\begin{pmatrix} 1 \\ y_0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, among many others.

If we have a tiling that is invariant under a translation v , it will also be invariant under multiples $2v$, $3v$, $-v$, etc. More generally, if a tiling is invariant under two translations v and w , it will also be invariant under the translations $av + bw$ where $a, b \in \mathbb{Z}$.

Example 3.2.2

1. The standard square tiling of the plane by unit squares is invariant under all translations of the form $v = \begin{pmatrix} x \\ y \end{pmatrix}$ with integers x and y .
2. The tiling by T-tetrominoes in **Figure 3.6** is invariant under translations by the yellow vectors (and many others).

**Figure 3.6** A periodic tiling by T-tetrominoes

Definition 3.2.2 Two translations v and w of the plane are called *independent* if they point in different directions. More precisely, this is the case if and only if the vector equation $av + bw = 0$ forces that $a = b = 0$. For two independent translations v and w , the set $\Lambda = \{av + bw : a, b \in \mathbb{Z}\}$ is called the *lattice generated by v and w* .

As we will be interested in periodic shapes and tilings in this chapter, the concept of a lattice is an important concept that allows us to describe the periodic behavior of the shape.

Note that when we scale a lattice by a factor or rotate a lattice about the origin, we usually obtain a different lattice. Here are two examples:

Example 3.2.3 The *square lattice* is the lattice of all translations by vectors with integral coordinates. It can be generated by $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $w = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. The same lattice can be generated in many other ways. For instance, also $v = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $w = \begin{pmatrix} 5 \\ 3 \end{pmatrix}$ generate the square lattice. All lattices obtained from this square lattice by scaling or rotating are still called square lattices.

Example 3.2.4 A *hexagonal lattice* can be generated by the edge vectors of an equilateral triangle. A particular example is generated by $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $w = \begin{pmatrix} 1/2 \\ \sqrt{3}/2 \end{pmatrix}$.

The following exercise gives another description of the hexagonal lattice and shows that it is sometimes easier to describe shapes in higher dimensions:

Exercise 3.2.3 Consider the plane E given by $x + y + z = 0$ in $\mathbb{R}^3$. Show that the set $\Lambda = \{(x, y, z) \in \mathbb{Z}^3 : x + y + z = 0\}$ is a hexagonal lattice in E . To do so, you need to find two elements of Λ that are edge vectors of an equilateral triangle and that also generate Λ .

Solution: Let $v = 1, 0, -1$ and $w = 0, 1, -1$. Then any vector $(x, y, z) \in \Lambda$ can be written as $xv + yw$, because this way the first two coordinates have the correct value, and the condition $x + y + z$ forces the third coordinate to be correct as well. Thus v and w generate Λ . To show that they are the edge vectors of an equilateral triangle, observe that they and their difference $v - w = (1, -1, 0)$ have the same length $\sqrt{2}$.

Suppose two lattices Λ_1 and Λ_2 are generated by vectors v_1, w_1 and v_2, w_2 , respectively. How do we find out whether the two lattices are equal? To do so, we need to check whether v_1 and w_1 are contained in Λ_2 and whether v_2 and w_2 are contained in Λ_1 .

For example, if the two lattices are generated by

$$\begin{aligned}\Lambda_1 : v_1 &= \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad w_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\ \Lambda_2 : v_2 &= \begin{pmatrix} -2 \\ 5 \end{pmatrix} \quad \text{and} \quad w_2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix},\end{aligned}$$

then we can write

$$v_2 = -2v_1 + 5w_1 \quad \text{and} \quad w_2 = v_1 - 3w_1$$

so that both v_2 and w_2 are contained in Λ_1 . Thus all of Λ_2 is contained in Λ_1 .

But we can also write

$$v_1 = -3v_2 - 5w_2 \quad \text{and} \quad w_1 = -v_2 - 2w_2$$

so that also all of Λ_1 is contained in Λ_2 , and the two lattices are equal. Note that in all these equations, the coefficients need to be integers.

Exercise 3.2.4 Three lattices Λ_1, Λ_2 , and Λ_3 are generated each as follows:

$$\begin{aligned}\Lambda_1 : v_1 &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad w_1 = \begin{pmatrix} -2 \\ 1 \end{pmatrix} \\ \Lambda_2 : v_2 &= \begin{pmatrix} -2 \\ 3 \end{pmatrix} \quad \text{and} \quad w_2 = \begin{pmatrix} 1 \\ -3 \end{pmatrix} \\ \Lambda_3 : v_3 &= \begin{pmatrix} -5 \\ 1 \end{pmatrix} \quad \text{and} \quad w_3 = \begin{pmatrix} -3 \\ 0 \end{pmatrix}\end{aligned}$$

Two of these lattices are equal to each other, while the third isn't. Determine which are equal, and explain why the third one isn't.

Solution: Λ_1 and Λ_3 are equal, because

$$v_3 = -v_1 + 2w_1 \quad \text{and} \quad w_3 = -v_1 + w_1$$

and

$$v_1 = v_3 - 2w_3 \quad \text{and} \quad w_1 = v_3 - w_3$$

However,

$$v_2 = \frac{4}{3}v_1 + \frac{5}{3}w_1 \quad \text{and} \quad w_2 = -\frac{5}{3}v_1 - \frac{4}{3}w_1,$$

so Λ_2 is not equal to Λ_1 .

If a lattice is generated by two vectors v and w , these vectors span a parallelogram that tiles the plane by using all translations from the lattice. This is a special case of an important concept:

Definition 3.2.5 A shape $S \subset \mathbb{R}^2$ is called a *fundamental shape* for a lattice Λ if the following two conditions are satisfied:

1. For every point $p \in \mathbb{R}^2$ there is a translation $v \in \Lambda$ such that $p + v \in S$.
2. If for a translation $v \in \Lambda$ and a point $p \in \mathbb{R}^2$ also the translate $p + v \in S$, then p lies in the boundary of S .

In other words: A fundamental shape is a single tile that tiles the plane, using precisely the translations in Λ . Evidently, the lattice parallelogram spanned by two generators of Λ is such a fundamental shape. There are, however, many others. One way to obtain them is by modifying the edges of the lattice parallelogram in such a way that opposite edges undergo the same modifications.

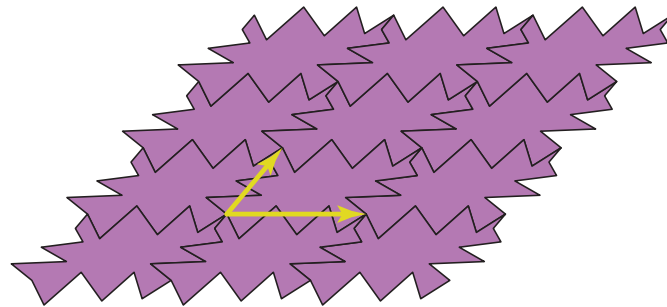


Figure 3.7 Fundamental shape of a lattice with two generating vectors

Exercise 3.2.6 Show that a hexagonal lattice has a fundamental region that has the shape of a regular hexagon.

Solution: This is best done the other way around: Take the periodic tiling of the plane by regular hexagons — it has as its fundamental domain one of the regular hexagons, and its period lattice is the hexagonal lattice.

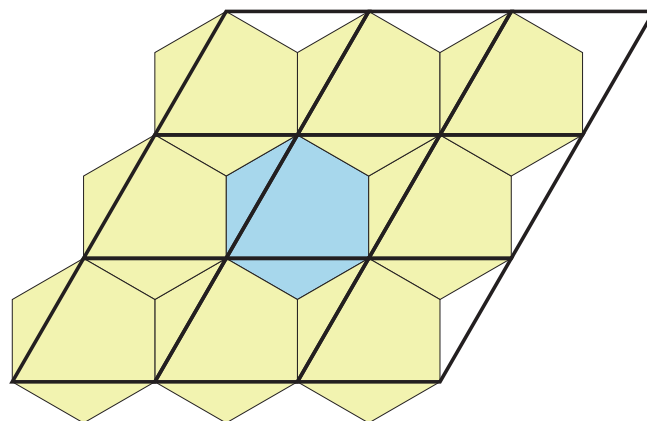


Figure 3.8 The hexagonal lattice

Exercise 3.2.7 A quadrilateral has its vertices at the points $(1, -1)$, $(2, 4)$, $(1, 3)$, and $(-2, 4)$, in counterclockwise order. By rotating the tile repeatedly about the midpoints of each edge, we obtain a periodic tiling of the plane. Determine a basis for its period lattice.

Solution: A fundamental domain of the periodic tiling is given by the quadrilateral together with one rotated copy, and the period lattice is spanned by $\begin{pmatrix} 4 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 4 \end{pmatrix}$.

In classifying tilings of the plane, we will distinguish three cases:

1. If a tiling is invariant under no translation at all, we call it non-periodic.
2. If the set of all translations that leave a tiling invariant is of the form $\{av : a \in \mathbb{Z}\}$, then the tiling is called *singly periodic*.
3. If the set of a translations that leave a tiling invariant is a lattice, i.e., of the form $\Lambda = \{av + bw : a, b \in \mathbb{Z}\}$ with v and w independent, then the tiling is called *doubly periodic* or just *periodic*. In this case, we call Λ the *lattice* of the tiling.

In the plane, these are all possibilities, because there are at most two independent directions.

Exercise 3.2.8 For each pentomino, find a periodic tiling of the plane with that pentomino, and determine two generating vectors for its lattice. Note that there can be many different solutions for each pentomino. Write the vectors in coordinates, assuming that the pentominoes are placed over a unit square lattice. For example, the periodic tiling by the T-tetromino in **Figure 3.6** has a lattice that can be generated by the vectors $\begin{pmatrix} 3 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -2 \\ 2 \end{pmatrix}$.

Exercise 3.2.9 For each hexiamond, find a periodic tiling of the plane with that hexiamond, and determine two generating vectors for its lattice.

Exercise 3.2.10 Find a polyomino that does not tile the plane periodically. Give a reason why this happens.

Exploration 3.2.5 We are going to find periodic tilings of the plane with rhombi. Recall that a rhombus is a parallelogram all of whose sides are of equal length. We ask that all rhombi in our tilings will have edge length 1, and tile edge-to-edge, i.e., two rhombi either share a vertex, or an entire edge, or are disjoint. There is the trivial example where a single rhombus tiles the plane by just translating it by its edge vectors. The lattice of the tiling is spanned by the edge vectors of the rhombus, and a single rhombus is a fundamental shape. Each corner is shared by four rhombi.

1. Find periodic tilings of the plane that use a single rhombus that are different from the trivial example.
2. What lattices can you get this way?
3. Find periodic tilings of the plane that use copies of two different rhombi.
4. Describe fundamental domains for your examples that consist entirely of rhombi.

5. Try to find examples that use a different number of rhombi of each kind in a fundamental domain.

6. Can you find examples where five, six, or more rhombi share a corner?

3.3 ARCHIMEDEAN TILINGS

In this section, we will discuss what tilings we can create by just using regular polygons. We need to be precise about how we allow the polygons to fit together, and we will be rather restrictive about this. Nevertheless, the classification process itself will produce many examples we probably wouldn't think about if we just tried to imagine the possibilities.

Definition 3.3.1 *A tiling is edge-to-edge if any two tiles either are disjoint, meet at a single corner, or have a single edge in common.*

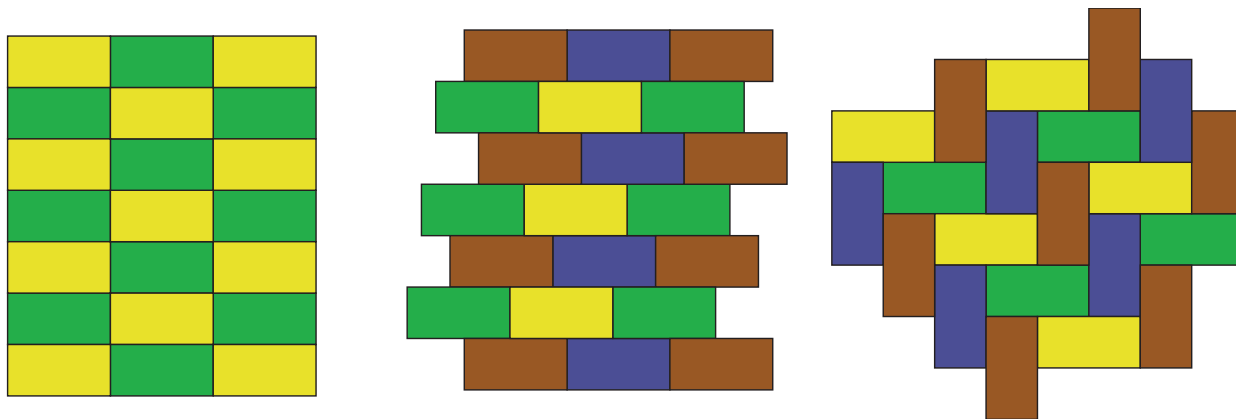


Figure 3.9 Some possible tilings with rectangles

Example 3.3.1 In Figure 3.9 we see three tilings by 2×1 rectangles. Only the first is edge-to-edge.

Exercise 3.3.2 *Find a hexagon that allows tiling the plane both edge-to-edge and not edge-to-edge.*

Solution: An idea is to modify the tiling by rectangles that are not edge-to-edge by adding vertices.

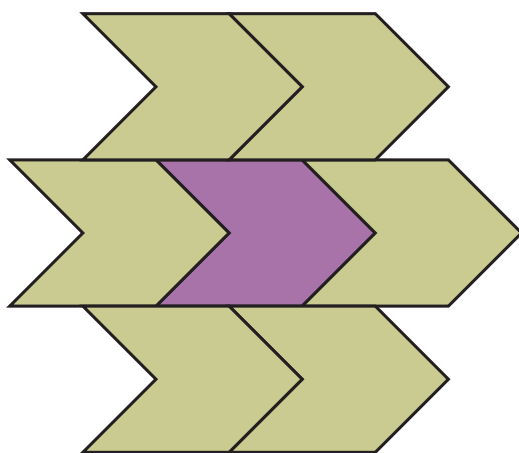


Figure 3.10 A hexagon that does not tile edge-to-edge

The most regular class of tilings conceivable allows only for regular polygons of one type as a tile, and requires that all tiles are arranged edge-to-edge. This class consists of only three tilings — namely, the ones by regular triangles, squares, and hexagons.

If we allow different kinds of regular polygons to occur in a tiling, we are asking for trouble: For instance, take the standard regular tiling by triangles. Six of the triangles around a corner fit together to a regular hexagon, so we can more or less randomly replace such six triangles with a hexagon, and create an abundance of such tilings without hope for a classification.

One way to reduce the number of possibilities is to impose more regularity:

vertex
transitive

Definition 3.3.3 A tiling is called *vertex transitive* if for any two vertices, there is a symmetry of the tiling that takes one vertex to the other.

All tilings in **Figure 3.9** are vertex transitive. In **Figure 3.11** you can see periodic tilings that are not vertex transitive. They are both supposed to extend periodically. The second one is edge-to-edge.

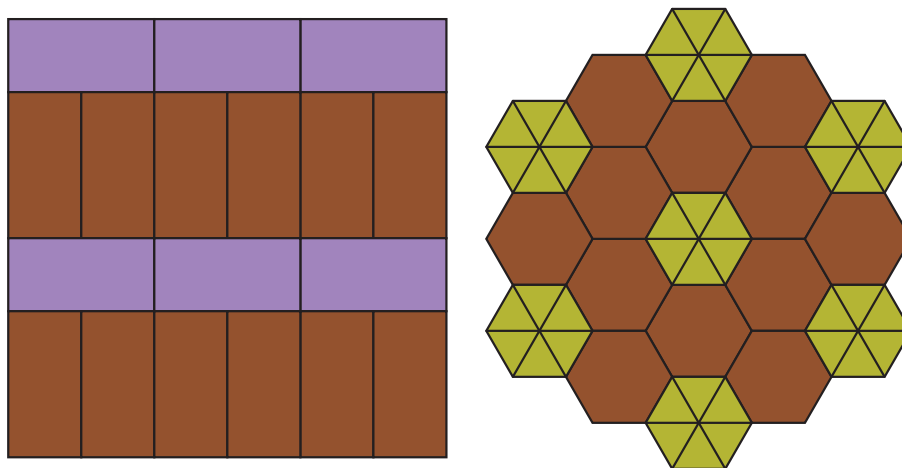


Figure 3.11 Two tilings that are not vertex transitive

Now we are ready to define Archimedean tilings. They were first investigated by JOHANNES KEPLER in his book *Harmonice Mundi* from 1619.

Archimedean

Definition 3.3.4 A tiling is called *Archimedean or uniform* if the following conditions are satisfied:

1. All tiles are regular polygons.
2. The tiling is edge-to-edge.
3. The tiling is vertex transitive.

The goal of this section will be to classify all Archimedean tilings. Classification is a typical scientific activity, and mathematicians indulge in it, too.

Ideally, coming up with a mathematical classification involves two steps:

1. Discover examples, and make a list.
2. Verify that all possible examples are in your list.

While the first step is common excitement to all scientists pursuing classifications, the second is the sole pleasure of the mathematician, as it involves logical reasoning and proofs. However, this idealistic description is rarely what

happens: Very often, an attempt to prove that a list of specimens is complete fails—and instead helps to find even more examples.

We first try to find examples of Archimedean tilings. In the simplest case, the tiling will consist of only one type of polygon. This can only happen for triangles, squares, and hexagons, because these are the only regular polygons that fit around a vertex without gaps. The three resulting tilings are also called *Platonic tilings*.

Are there others? In **Figure 3.12**, you can see two ways to tile the plane periodically with triangles and squares, the first one is Archimedean, the second one is not.

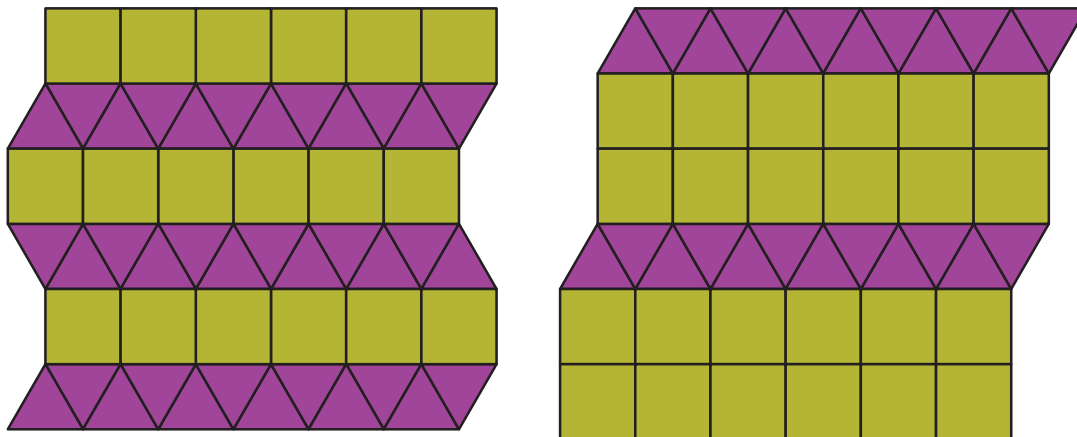


Figure 3.12 Archimedean vs Non-Archimedean

How would we go about finding *all* Archimedean tilings? A good approach is to first find out which different regular polygons can fit around a common corner.

It is clear that we will need at least three polygons, and it is easy to see that there can be at most six: That's because every polygon will contribute to the total angle 360° at the corner, and the smallest contribution is that of an equilateral triangle, i.e. 60° .

Hence, we will proceed by a *case distinction*: We will look at tilings where 3, 4, 5, or 6 tiles surround a common corner. By the condition that all corners look alike, each case will apply to all corners of the tiling.

CASE 1: THREE POLYGONS AROUND A CORNER

Let's begin by studying all possible Archimedean tilings with *valency* 3, i.e., where there are three polygons around each corner. There are surprisingly many possibilities to fit three regular polygons around a corner. For instance, did you know you can do it with a triangle, an octagon, and an icositetragon (a 24-gon, that is)?

Here is maybe a good place to recall how to compute the interior angle of a regular n -gon:

Lemma 3.3.5 *The interior angle of a regular n -gon at a vertex is $\pi - 2\pi/n$.*

Proof. To see this, we look instead at the *exterior angle*: This angle describes by how much we need to turn left at a vertex when we walk counterclockwise around the polygon. After walking around the n -gon, we have completed a full turn, so that each exterior angle contributes $2\pi/n$ to this full turn. However, the sum of interior and exterior angles equals π , which provides us with the formula. $\square$

So, suppose we have an a -gon, b -gon, and c -gon huddling together around a vertex. Then all their interior angles need to add up to 2π . By our formula, this means that we need to have

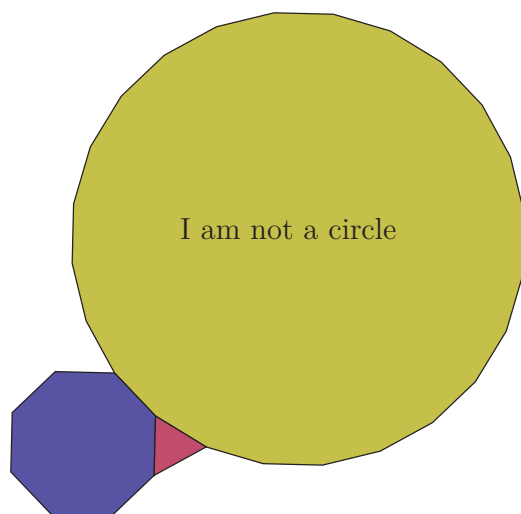


Figure 3.13 Regular icositetragon, octagon, and triangle fit around a vertex

$$\pi - \frac{2\pi}{a} + \pi - \frac{2\pi}{b} + \pi - \frac{2\pi}{c} = 2\pi$$

which can be simplified to

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{1}{2}.$$

It is easy to verify now that

$$\frac{1}{24} + \frac{1}{8} + \frac{1}{3} = \frac{1}{2},$$

verifying the claim of **Figure 3.13**.

Finding *all* solutions to this equation is not so easy; therefore we make this an exercise:

Exercise 3.3.6 Find all 10 solutions (up to permuting a, b, c) to the diophantine equation $\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{1}{2}$ in integers $a, b, c \geq 3$.

Solution: Let's assume $3 \leq a \leq b \leq c$. We have one solution $a = b = c = 6$. If we exclude that, we have necessarily $a \leq 5$ and $c \geq 7$. We distinguish three cases:

1. If $a = 3$, then $\frac{1}{b} + \frac{1}{c} = \frac{1}{6}$, or $c = 6 + \frac{36}{b-6}$. This forces $b - 6$ to be a divisor of 36, leading to the solutions $(3, 7, 42), (3, 8, 24), (3, 9, 18), (3, 12, 15), (3, 12, 12)$.
2. If $a = 4$, then $\frac{1}{b} + \frac{1}{c} = \frac{1}{4}$, or $c = 4 + \frac{16}{b-4}$. This forces $b - 4$ to be a divisor of 16, leading to the solutions $(4, 5, 20), (4, 6, 12), (4, 8, 8)$.
3. If $a = 5$, then $\frac{1}{b} + \frac{1}{c} = \frac{3}{10}$, or $3c = 10 + \frac{100}{3b-10}$. This forces $3b - 10$ to be a divisor of 100, leading to the only new solution $(5, 5, 10)$.

Fortunately, we do not really need to figure all this out, because most of the possibilities do not actually lead to an Archimedean tiling. To see why, let's suppose there is a polygon A in our tiling with an odd number a of edges. There

are two further polygons involved, say polygon B with b and polygon C with c edges. Let's follow the polygon A with a edges around its perimeter and see what polygons it neighbors on. At each vertex, we have polygons with a , b , and c edges. Thus, if one neighbor of A is B , the next neighbor must be C , the next B again, etc, until we complete one turn after an odd number of steps. This is only possible if B and C are the same polygons, i.e., if $b = c$.

So we divide our case 1 into two subcases: In the first case, we assume that a is odd and thus $b = c$, and in the second case, we will assume that a, b, c are all even. So, first, let a be odd, and thus $b = c$. Solving the angle condition

$$\frac{1}{a} + \frac{2}{b} = \frac{1}{2}$$

for b gives

$$b = \frac{4a}{a-2} = 4 + \frac{8}{a-2}.$$

Note the trick to use partial fractions at the end of the formula. Because b needs to be an integer, this helps to see now that $a - 2$ must be a divisor of 8, and odd as well. This means that necessarily $a = 3$ and $b = 12$. Trying this out shows that we can indeed extend this tiling as in **Figure 3.14**. Here, our quest shows a first success: We found indeed an Archimedean tiling!

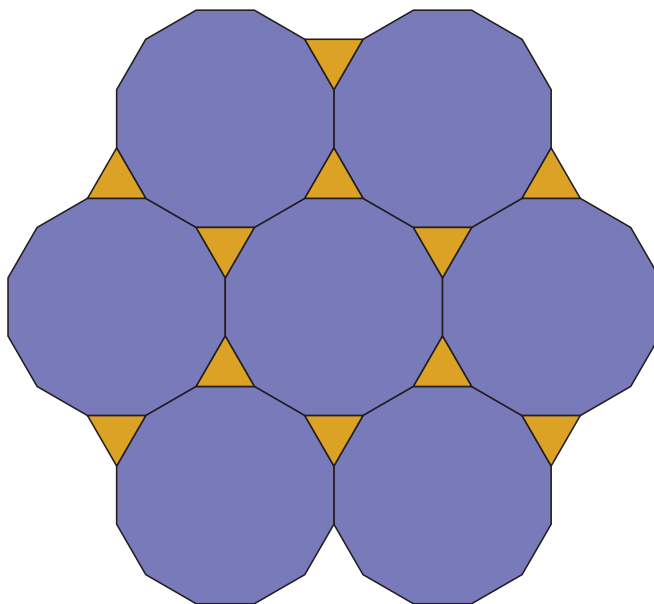


Figure 3.14 Two dodecagons and a triangle go Archimedes

The other subcase allows only polygons that have an even number of vertices. Writing $a' = 2a$, $b' = 2b$ and $c' = 2c$, our angle condition becomes

$$\frac{1}{a'} + \frac{1}{b'} + \frac{1}{c'} = 1$$

Note the special solution $a' = b' = c' = 3$, corresponding to the hexagon tiling. Any further solution must have one variable less than 3, because we cannot increase them all and keep the equation valid. As 2 is the smallest possible value, we can assume $a' = 2$, and end up with the equation

$$\frac{1}{b'} + \frac{1}{c'} = \frac{1}{2}$$

Solving for c' gives

$$c' = 2 + \frac{4}{b' - 2}.$$

As c' needs to be an integer greater or equal to one, $b' - 2$ needs to be a divisor of 4. This leaves us with

$$(b', c') \in \{(3, 6), (4, 4), (6, 3)\}$$

and give us triples of possible polygons with $(a, b, c) \in \{(4, 6, 12), (4, 8, 8)\}$. Both are new candidates for Archimedean tilings, and it is easy to see that there is only one way to extend the tilings so that all corners are congruent.

CASE 2: FOUR POLYGONS AROUND A CORNER

We now turn our attention to the possibility that we have always four polygons around each vertex.

Having higher valency seems to be harder. However, there are also simplifications. The simplest case with valency 4 is that of having four squares around a corner. Thus any *other* solution will have to have a triangle, again because we can't decrease all angles.

Exercise 3.3.7 Find all quartets of regular polygons that fit around a common corner.

Solution: The angle condition means

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} = 1.$$

Excluding $a = b = c = d = 4$, we can assume $a = 3$. If also another number (say b) equals 3, we get

$$\frac{1}{c} + \frac{1}{d} = \frac{1}{3}.$$

Solving for d gives

$$d = 3 + \frac{9}{c - 3}$$

which can be solved in integers by $(c, d) \in \{(4, 12), (6, 6), (12, 4)\}$.

If no other number equals five, two of them must be 4, because otherwise

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d} \geq \frac{1}{3} + \frac{1}{4} + \frac{3}{5} > 1.$$

So we can assume that $b = c = 4$ which leads to $d = 6$. Hence, all solutions are from the list

$$(3, 3, 4, 12), (3, 3, 6, 6), (3, 4, 4, 6), (4, 4, 4, 4)$$

Now we will do again a case distinction into subcases, counting how many triangles do occur. We can't have more than three, because then we had more than four polygons around a vertex, which will be discussed in the next case. Hence we need to consider the subcases of having none, one, two, or three triangles around a vertex. Having no triangle means that we are in the case $(4, 4, 4, 4)$ of four squares, as discussed.

In the next subcase, we assume that there is only one triangle. Then we have, around any vertex, a triangle, a b -gon, c -gon, and d -gon (in this order) with $a, b, c \geq 4$, and $d = 3$.

Using the same argument as in the previous case, we follow the triangle around its perimeter and observe that adjacent to its sides are alternately a b -gon and a d -gon, so, we need to have $b = d$. Thus the angle condition becomes

$$\frac{1}{3} + \frac{2}{b} + \frac{1}{c} + \frac{1}{c} = 1.$$

Solving this for b gives

$$b = 3 + \frac{9}{2c - 3}.$$

As b is an integer, we need $2c - 3 \in \{1, 3, 9\}$, or $c \in \{2, 3, 9\}$. The first solution is too small, the second forbidden because we are only looking for one triangle. Thus we have one more candidate for Archimedean tilings, with vertex type 3-4-6-4. Note that the possibility 3-4-4-6 is excluded, because the neighbors of the triangle have to be congruent.

Exercise 3.3.8 Find a periodic tiling of the plane that is edge-to-edge, uses only regular polygons, and where the only vertex types allowed are 3-4-6-4 and 3-4-4-6 (or 3-6-4-4).

Solution: See Figure 3.15 for a tiling using only vertex type 3-4-4-6.

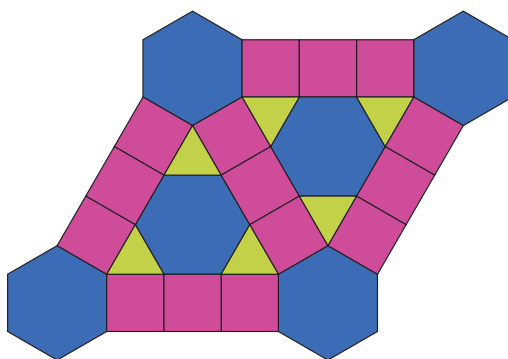


Figure 3.15 Vertex type 3-4-4-6

The second subcase asks for having precisely two triangles. We now disregard the order of the polygons around the vertex so that we can assume $a = b = 3$. The angle condition becomes

$$\frac{1}{3} + \frac{1}{3} + \frac{1}{c} + \frac{1}{d} = 1.$$

Solving for d gives

$$d = 3 + \frac{9}{c - 3}.$$

As d is an integer, $c - 3$ divides 9, so that $c \in \{4, 6, 12\}$, and $d \in \{12, 6, 4\}$. This leaves us with the possible polygon quartets $(3, 3, 4, 12)$ and $(3, 3, 6, 6)$. The first case is impossible in an Archimedean tiling: If the vertex type is 3-4-3-12 so that the two triangles are opposite, we get a contradiction because then squares and 12-gons have to alternate along the edges of a triangle. If the vertex type is 3-3-4-12, one of the two triangles has a square opposite across its corners. Then the 'other' triangle and the 12-gon have to alternate along the edges of the first triangle, which again is impossible.

Exercise 3.3.9 Find a periodic tiling of the plane that is edge-to-edge and uses regular triangles, squares, hexagons, and 12-gons as polygons. All these polygons have to appear, and no others.

Solution: See Figure 3.16.

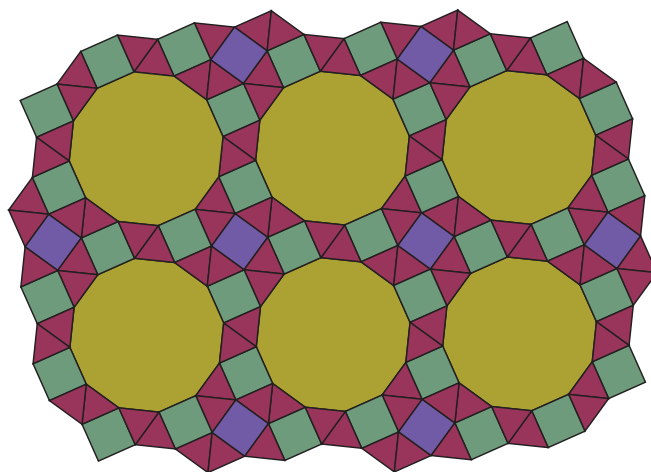


Figure 3.16 Triangle, square, hexagon, and dodecagon

Finally, in the last subcase, having three triangles is impossible, as the remaining polygon would have to have a 180° angle.

CASE 3: FIVE POLYGONS AROUND A VERTEX

Now we deal with the case of valency 5. If only two or even fewer polygons are triangles, the angle sum would be larger than 2π . Thus we have $a = b = c = 3$, and the angle condition becomes

$$\frac{1}{3} + \frac{1}{3} + \frac{1}{3} + \frac{1}{d} + \frac{1}{e} = \frac{3}{2}$$

which, solved for e , means

$$e = 2 + \frac{4}{d-2}.$$

This has the integral solutions $(d, e) = (3, 6)$ and $(d, e) = (4, 4)$.

This allows for the following possibilities for (a, b, c, d, e) : $(3, 3, 3, 3, 6)$, $(3, 3, 3, 4, 4)$, $(3, 3, 4, 3, 4)$.

Now that we have listed all possible stars that might occur in an Archimedean tiling, we need to check which can actually be realized. Surprisingly, all cases can actually be realized. Note that sometimes it is possible to have the same polygons around a vertex, but in a different order. The labels in the figure denote the number of edges of the n -gons around a vertex of the tiling. Observe that there are two different tilings which are labeled 3-3-3-3-6. They look very much the same, but they are not quite congruent: One is only the mirror image of the other.

To conclude our investigation:

Theorem 3.3.10 *All Archimedean tilings are given by the above list.*

As a consequence of the classification we obtain

Corollary 3.3.11 *All Archimedean tilings are periodic.*

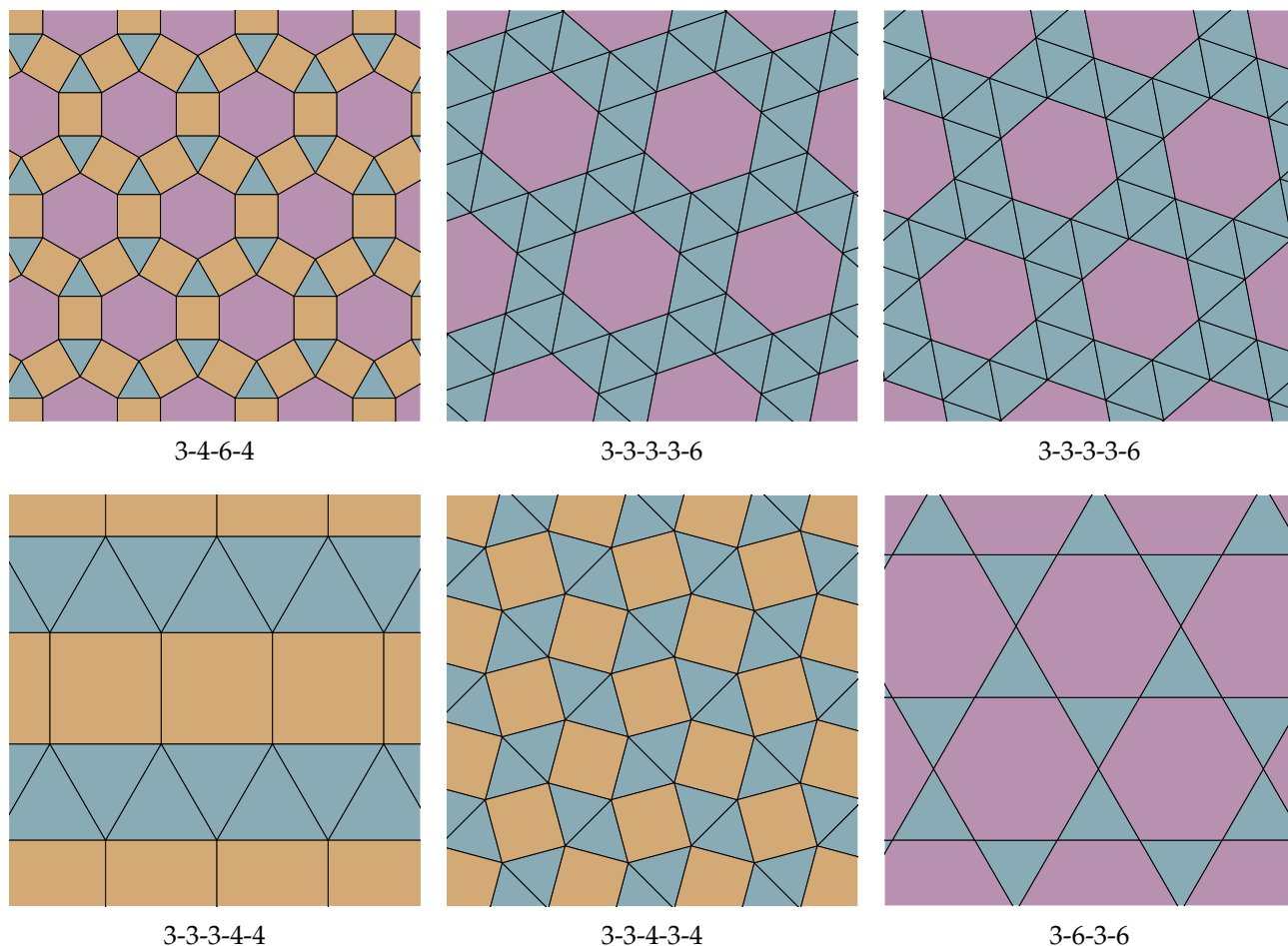


Figure 3.17 The Archimedean Tilings I

Exercise 3.3.12 Suppose a regular 12-gon has circumradius 1, i.e., it can be inscribed in a unit circle. Find its area. You are not allowed to use trigonometry. Instead, play around with Figure 3.19 — the answer is simpler than you would expect.

Exploration 3.3.2 We are going to explore periodic edge-to-edge tilings of the plane that consist just of squares and equilateral triangles whose edges have unit length. In Figure 3.21 you can see all four different ways in which these tiles can be grouped around a single corner. In this section, we have classified all tilings where only one type may occur. Your goal is to find periodic tilings with two different types of vertices. For each of the six possible pairs of types, you can pursue the following

1. Find examples of periodic tilings.
2. Explain why maybe there isn't any.
3. Determine a basis of the period lattice.
4. Find a pretty fundamental domain.
5. Classify all possibilities.

A particular fertile case is illustrated in Figure 3.20. Can you find others like this, and maybe classify them?

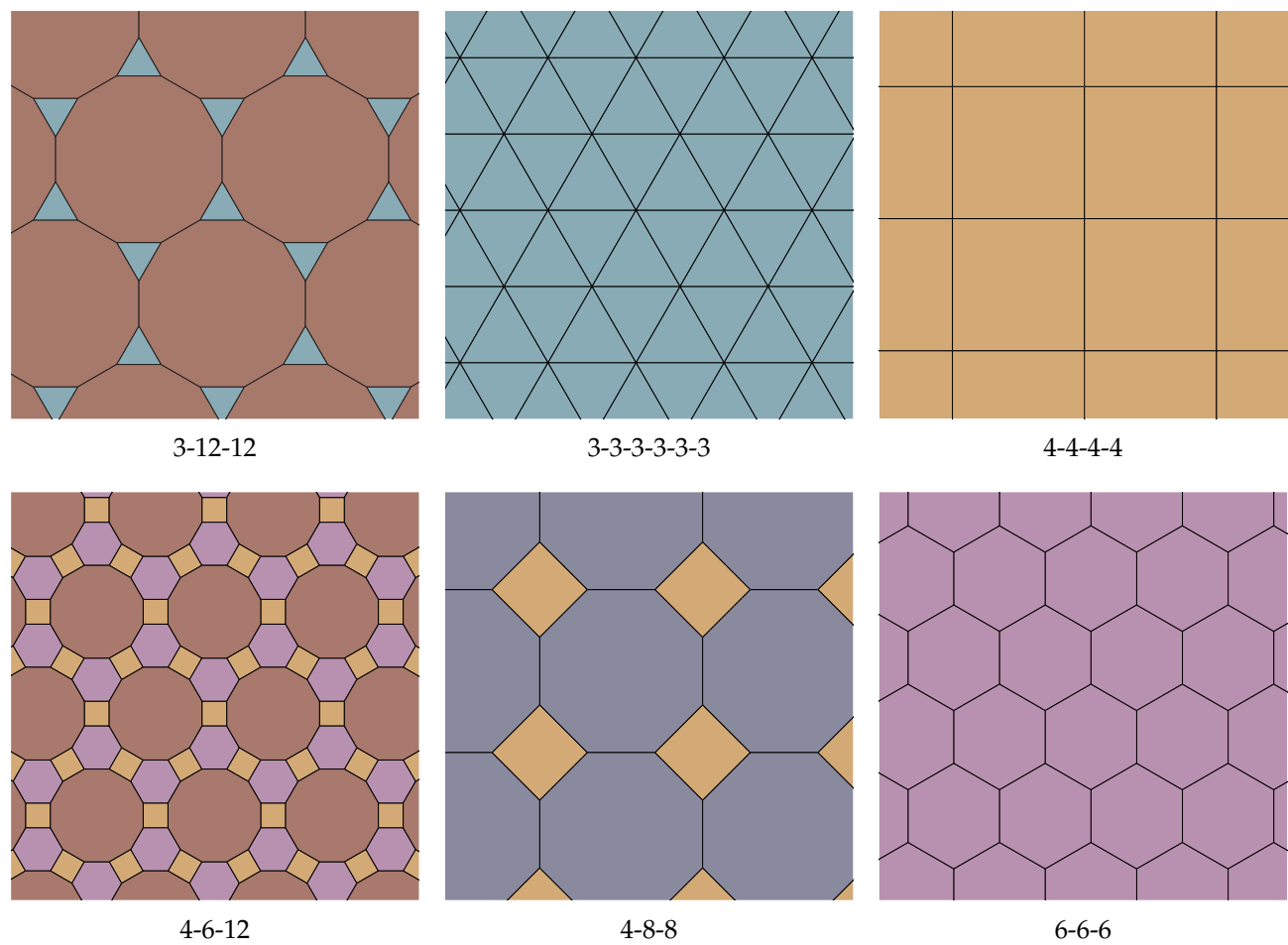


Figure 3.18 The Archimedean Tilings II

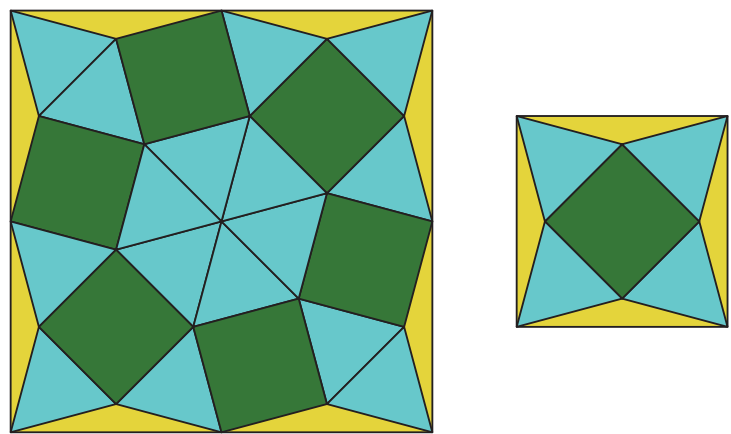


Figure 3.19 The magic disappearance of the yellow triangles

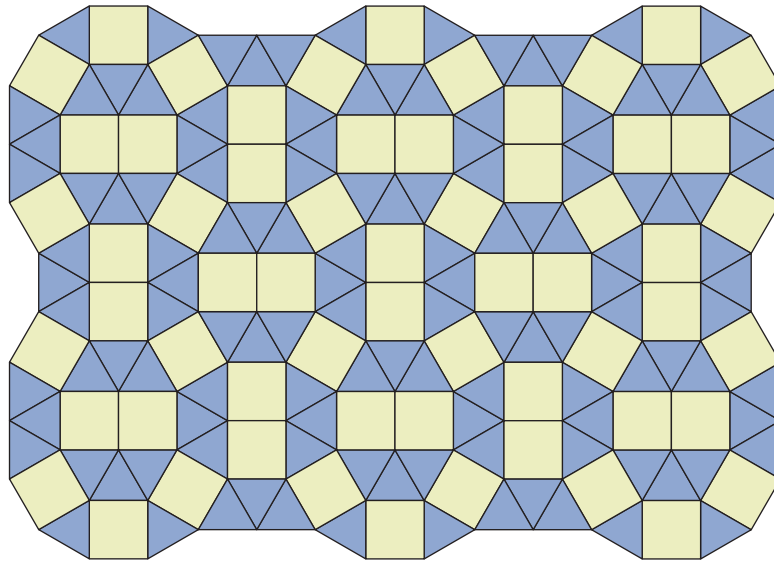


Figure 3.20 Using two vertex types

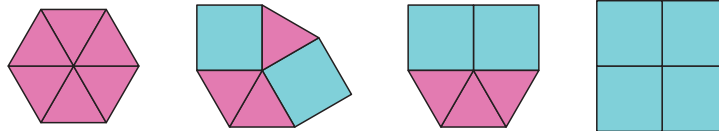


Figure 3.21 How to be surrounded by triangles and squares

3.4 WHICH SHAPES TILE?

How do we determine whether a given shape tiles the entire Euclidean plane? This is not just a difficult problem — it is in fact *unsolvable*. More precisely, there is no computer program that can determine for an arbitrary set of polygons whether it can be used to tile the plane or not. This doesn't mean that the question isn't interesting anymore. There are many cases where we have an answer to the question, and we will discuss a few of them.

Definition 3.4.1 We call a shape simple if its boundary is a single closed curve without self-intersection.

Example 3.4.1 A square is a simple shape, but the union of two squares that meet at a single corner is not. Also, a square with a hole in the center is not simple.

It will be convenient to divide the boundary of a simple shape into segments. This leads to the definition:

Definition 3.4.2 A generalized polygon is a simple shape with distinguished points $P_1, \dots, P_n$ in counterclockwise order on its boundary. We call the P_i vertices and the segments on the boundary from P_i to P_{i+1} edges.

Example 3.4.2

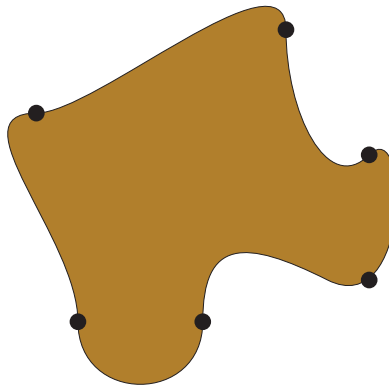


Figure 3.22 A generalized hexagon

Definition 3.4.3 Two edges P_iP_{i+1} and P_jP_{j+1} of a generalized polygon are called parallel if there is a translation that takes the edge P_iP_{i+1} to the edge P_jP_{j+1} so that P_i is taken to P_{j+1} and P_{i+1} to P_j . A generalized parallelogram is a generalized 4-gon in which both pairs of opposite edges are parallel.

Here is our first tiling result:

Theorem 3.4.4 Any generalized parallelogram tiles the plane, just using translations.

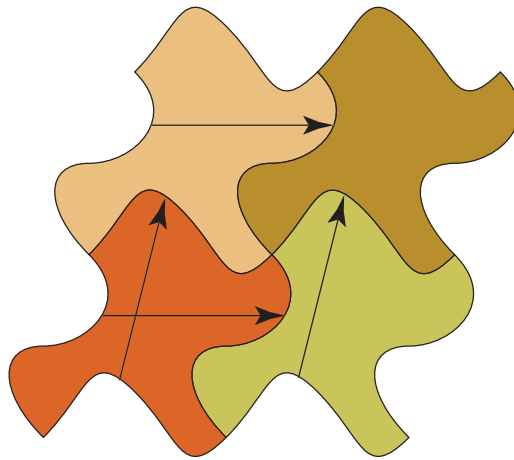


Figure 3.23 A generalized parallelogram tiles

Proof. We apply the translations that take an edge to its opposite edge to our given generalized parallelogram. Doing so, as in **Figure 3.23**, provides us with four copies of our shape that fit together along the edges by the definition of parallel. We also see that just four copies of our shape fit around a corner, adding up to a single full turn and thus fitting together there, too. Thus, we can extend the tiling to the entire plane, using the translations from the lattice generated by the edge translations. $\square$

Even though we have seen that regular hexagons tile the plane by translation, it might come as a surprise that the above theorem generalizes to hexagons:

Theorem 3.4.5 *Any generalized hexagon whose opposite edges are parallel tiles the plane, just using translations.*

Proof.

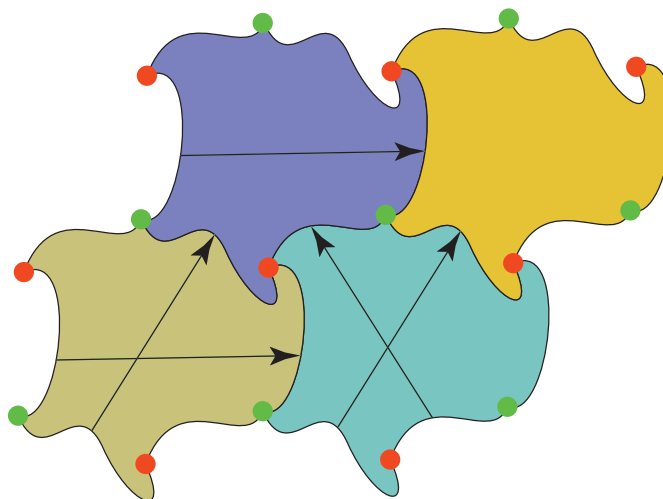


Figure 3.24 A parallelo-hexagon tiles, too

We again apply the translations that take an edge to its opposite edge to our given generalized hexagon. Again, the tiles fit together along the edges, as seen in **Figure 3.24**. This time, the six corners of the hexagon are matched up in

two triplets, around each of which we have a complete turn, so that again the hexagons also fit around the corners together. $\square$

Exercise 3.4.6 *What goes wrong if we want to use the same argument to show that a generalized octagon where all opposite edges are parallel tiles the plane?*

Solution: This time, all eight corners have been brought together to a single vertex, making two full turns and thus creating an overlap.

These two criteria are somewhat limited: first, we rarely have shapes with many parallel edges, and second, we seldom expect our shapes to tile simply by translation.

Fortunately, there is a surprisingly useful condition due to JOHN CONWAY that also allows 180° rotations:

Definition 3.4.7 *We say that a generalized hexagon satisfies the Conway condition if its boundary edges $A, B, C, D, E,$ and F satisfy that*

1. *edges A and D are congruent by a translation,*
2. *arcs $B, C, E,$ and F are symmetric with respect to a 180° rotation about their respective centers.*

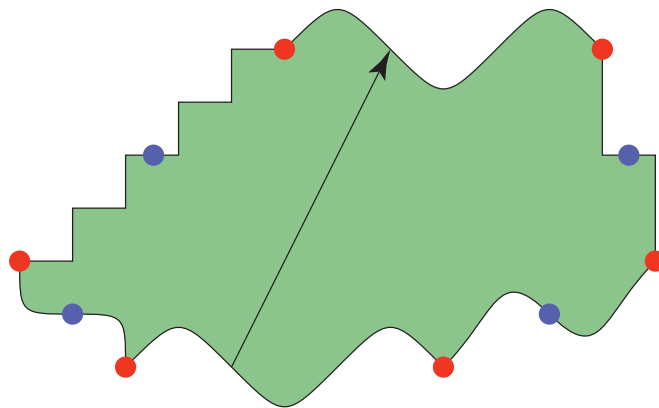


Figure 3.25 A hexagon satisfying the Conway condition

Example 3.4.3 In **Figure 3.25** the vertices are marked with red dots while the centers about which the edges are symmetric are marked with blue dots. The arrow indicates the two parallel edges.

Theorem 3.4.8 *If a generalized hexagon satisfies Conway's condition, it tiles the plane using translations and 180° rotations.*

Proof. We first use the translation we have to create a green strip of copies of our shape that fit together along their edges A and D . This strip has two boundary curves: One consists of the edges B and C repeating in the pattern $\dots BCBC \dots$, while the other consists of the edges E and F repeating in the pattern $\dots EFEF \dots$

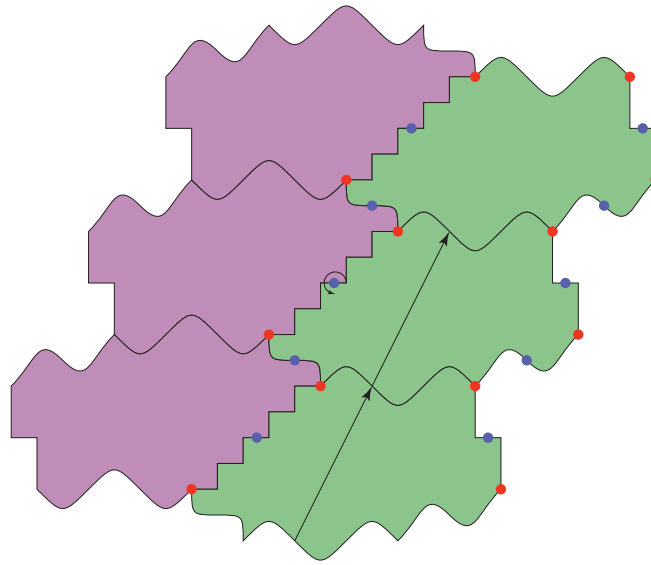


Figure 3.26 Tiling with Conway hexagons

Let's focus on the $\dots EFEF \dots$ curve. Each segment E or F is central symmetric with respect to their respective midpoints. Therefore, if we apply one of these 180° rotations to the entire $\dots EFEF \dots$ curve, it will match up with itself everywhere! Thus, if we rotate the strip shape about one of these centers, the image will be a congruent purple strip fitting together with the original strip along the $\dots EFEF \dots$ curve. Moreover, the *other* edge of the rotated strip will be a translational copy of the $\dots BCBC \dots$ curve of the green strip, so that we can indeed tile the entire plane now by replicating the two copies of the strip. $\square$

Remark 3.4.4 While the arguments in the proofs are essentially correct, we were a little vague about the reason why the shapes fit together at the vertices. This can be fixed with an argument from topology involving simple connectivity and covering spaces. For the concrete shapes for which we are going to use this theorem, this will be as evident as the figure.

Example 3.4.5 To apply the Conway criterion, we need to subdivide the boundary of a simple shape into six edges. This can be somewhat counterintuitive as in **Figure 3.27**, where we have marked the vertices of the generalized hexagon with really fat red dots.

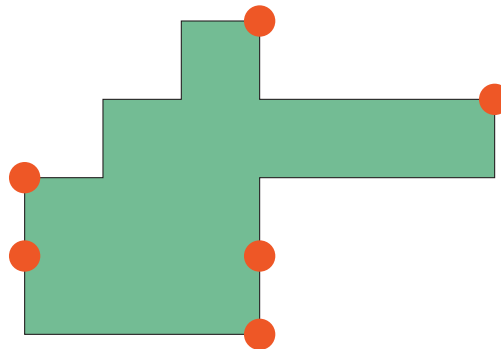


Figure 3.27 A tiling dodecomino

Exercise 3.4.9 All hexominoes tile the plane. Show this for ten of them of your choice, using the Conway condition.

3.5 DISSECTIONS

Dissection problems usually ask to cut one figure into pieces and then to reassemble the pieces into another figure.

While problems of this type clearly belong to mathematical pastimes, there are a few important applications within and outside mathematics. For instance, it is an open problem to find optimal or near-optimal layout patterns for tailoring pieces of a garment so that as little cloth as possible is wasted when the pieces are cut.

We will address the following

Problem 3.5.1 *Given two Euclidean polygons, find a decomposition of them into as few as possible congruent polygonal pieces.*

For this to be possible, the two polygons must necessarily have the same area. Surprisingly, this is also sufficient:

Theorem 3.5.2 *Any two Euclidean polygons of equal area can be dissected into congruent polygonal pieces.*

To see this, we first look at a very special case:

Proposition 3.5.3 *Any rectangle can be dissected into a square of the same area.*

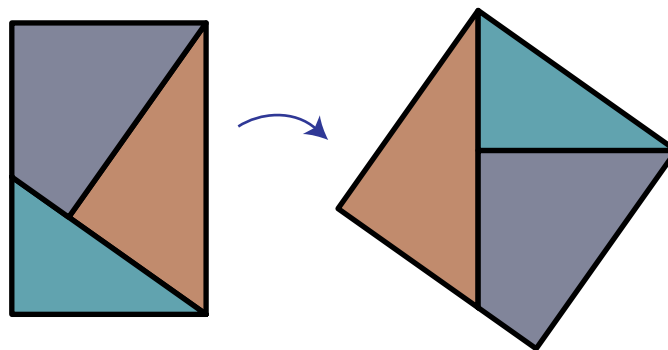


Figure 3.28 Dissecting a rectangle into a square

Proof. First, we ensure that the rectangle is well-proportioned. i.e., it is not more than twice as long as thin. Otherwise, divide the rectangle parallel to the shorter edge into two congruent rectangles, and reassemble them. Repeat this until the size condition is satisfied. To a well-proportioned rectangle, we can apply the dissection of [Figure 3.1](#).

This needs some explanation: Let the shorter edge of the rectangle have length a and the longer length b so that the corners are at the coordinates $(0,0)$, $(a,0)$, (a,b) and $(0,b)$. Then the diagonal at $(a,0)$ is constructed to be of length $\sqrt{ab}$ so that it cuts the other edge at $(0, \sqrt{a(b-a)})$. The diagonal at (a,b) is perpendicular to the first diagonal and intersects it at $((a - \sqrt{a(b-a)}, b - a)$. The pieces can be reassembled by translating them into a dissection of a square with edge length $\sqrt{ab}$, as shown. $\square$

Exercise 3.5.4 *Show that the pieces fit together.*

Corollary 3.5.5 *Any two rectangles of the same area can be dissected into each other.*

Proof. Using [Proposition 3.5.3](#), we can dissect both rectangles in the same square. Thus, cutting the pieces of the square form the first dissection along the lines of the second dissection will produce a much larger dissection of the square, which can be used to assemble either rectangle. $\square$

We will now show that *every* polygon (and not just a rectangle) can be dissected into a square of the same area.

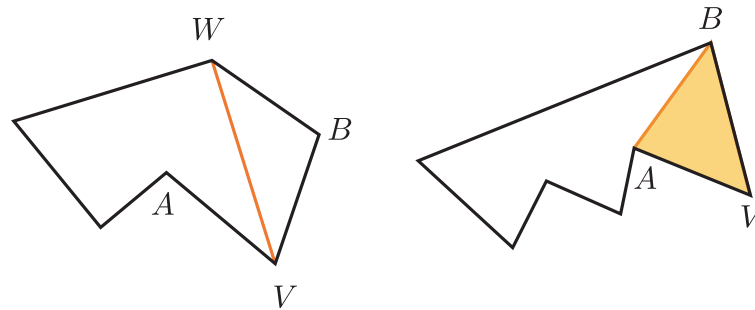


Figure 3.29 Finding a secant in a polygon

First, we decompose the polygon into triangles:

Lemma 3.5.6 *Any polygon has a secant, i.e., a line segment that connects two vertices and remains in the interior of the polygon.*

Proof. To find this secant, pretend that the polygon is actually the floor plan of a room, and we are standing at one vertex V . The two adjacent walls lead to two vertices A and B , which we can see. If we can see yet another vertex W from our position, we have found our secant VW . If we can't see another vertex, nothing obstructs our view in the triangular region formed by A , B , and V , and thus A and B can be joined by a secant. $\square$

Given such a secant, we can divide the polygon along it into two polygons with fewer vertices each, and treat them separately, finding new secants in each of them, until everything is divided into triangles.

As a further simplification, we cut all triangles into two pieces along one of their heights so that all triangles become right triangles.

Now we have a collection of right triangles, which will need to be dissected into a single square.

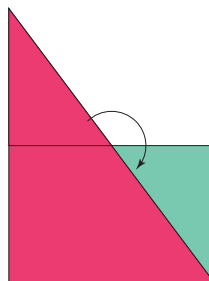


Figure 3.30 Dissecting a triangle into a rectangle

To do so, we dissect each right triangle into a rectangle. This can be done by dissecting the triangle into two pieces along a segment parallel to one of the legs and dividing the other leg into equal parts.

This leaves us with a collection of rectangles that most likely have different dimensions. So we choose a positive real a number smaller than all the edge lengths of all the rectangles. By [Corollary 3.5.5](#), each of them can be dissected into a rectangle of equal area with one edge having length a .

Then, the new rectangles can be lined edge to edge along their sides of length a up to form one very long rectangle with the short edge having length a .

Finally, this rectangle can be dissected into a square of the same area by [Proposition 3.5.3](#).

Exercise 3.5.7 Find a dissection of a 3×1 rectangle into a square, using as few tiles as possible.

Solution: The construction via well-proportioned rectangles yields a dissection with five pieces. Below is one that only needs three, found by Tessa Pilachowski:

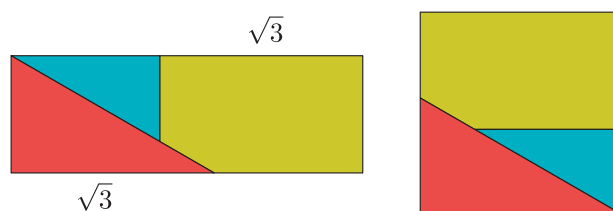


Figure 3.31 A 3×1 -rectangle dissected into a square

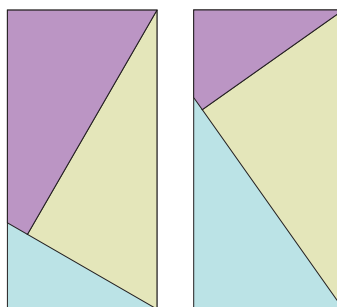


Figure 3.32

Exercise 3.5.8 A rectangle with corners at $(0, 0)$, $(1, 0)$, $(1, 2)$, and $(0, 2)$ needs to be dissected into a rectangle with proportions $2 : 3$. In [Figure 3.32](#), you can see a sketch of two such dissections, both very similar to the one we discussed for dissecting a well-proportioned rectangle into a square. Find the coordinates of the corner within the left rectangle so that this dissection actually works. Hint: First determine the length of the hypotenuse of the blue triangle. Then you can compute the coordinates of the vertices of the blue triangle, and the slope of its hypotenuse. The hypotenuse of the mauve triangle is orthogonal to that. Then use the point-slope form of lines to find the intersection.

Solution: The dimensions of the new rectangle are $2/\sqrt{3}$ and $\sqrt{3}$. The hypotenuse of the lower triangle must be one of these lengths, and the sketch suggests it is the shorter one. This allows to compute the slope of the two perpendicular line segments in the rectangle. Using the point slope form of lines, we determine the coordinates of the intersection as $x = 1 - \sqrt{3}/2$ and $y = 1/2$.

There is a second dissection when you assume that the hypotenuse is in fact the longer edge. Then the coordinates are $x = 1 - 2\sqrt{2}/3$ and $y = 4/3$.

Remark 3.5.1 The same theorem turns out to be false in Euclidean space. In 1910, DAVID HILBERT posed 27 mathematical problems (the famous Hilbert problems). Some of them are still unsolved today. One of them asked whether two polyhedra of equal volume in space could be dissected into congruent pieces. In 1920, MAX DEHN showed that this was impossible, even for such simple polyhedra like the regular tetrahedron and the cube.

Remark 3.5.2 If one drops the limitation that the cuts are along planes so that all pieces are again polyhedra, but allows for arbitrary sets, even more troublesome things happen: The *Banach-Tarski paradoxon* decomposes a ball of radius 1 into five subsets which can be reassembled by Euclidean congruences into two balls, *both* of radius 1! This result is extremely counterintuitive, because we know that volume must be preserved under dissections. However, the pieces used in the paradoxical dissection do not have a volume in any mathematical sense. The existence of such sets without volume follows from the *axiom of choice*, a widely used but often discussed axiom from set theory. It allows one to choose simultaneously elements from any collection of sets. If the number of sets is very large, it is generally impossible to make this selection constructively.

So we have found a procedure to dissect any polygon into another of the same area. This is *effective* in the sense that we can actually do it, if forced to. On the other hand, the number of pieces needed will be rather large in general, rendering the method useless both for practical purposes and for solving dissection puzzles.

There is no known general method that would provide optimal solutions. There are, however, a few methods available that apply in particular cases and give very good results. We will discuss one such method.

Suppose the two polygons (or, more generally, shapes) we want to dissect both tile the plane periodically. Assume in addition that the two tilings are invariant under the same lattice of translations. By overlaying the two tilings, the polygons of one tiling dissect the polygons of the other tiling into pieces that can be reassembled into polygons of the first kind.

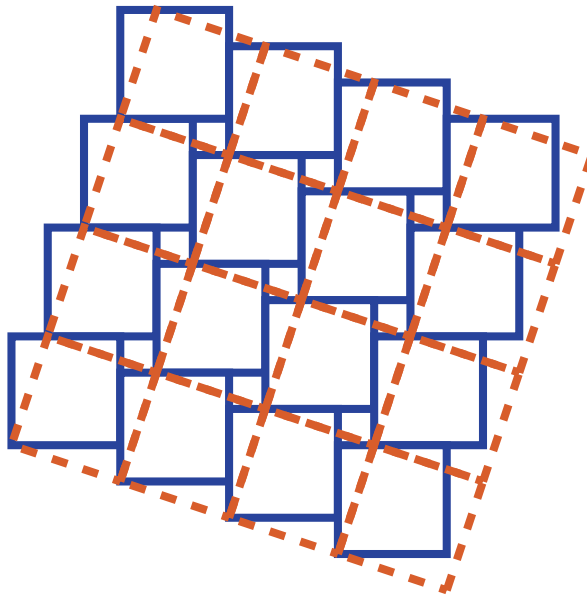


Figure 3.33 Big square into two smaller squares

Example 3.5.3 In **Figure 3.33** we illustrate how to dissect one big square into two smaller squares. We use two tilings, one just by big squares, and the other by the two smaller squares.

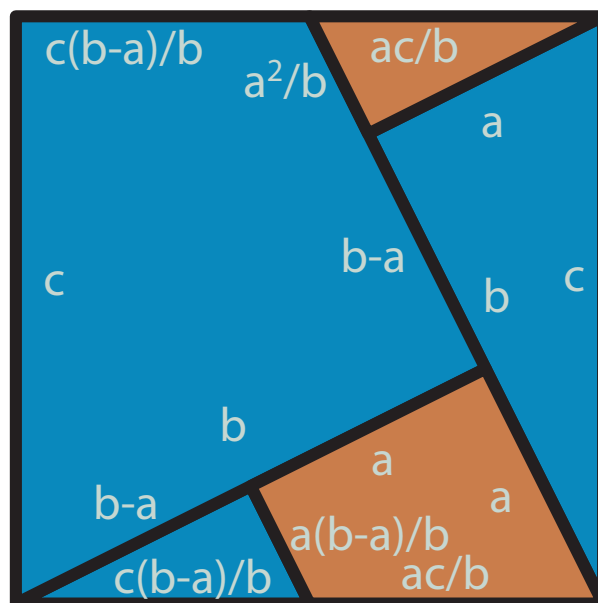


Figure 3.34 Pythagoras, dissected

As the area of the two small squares equals the area of the big square, we might suspect that Pythagoras' theorem is around the corner. We can see this by indicating the edge lengths of the pieces as in **Figure 3.34**.

Example 3.5.4 The *Swiss cross* is a cross consisting of five equal squares. As an illustration of the tiling method to find dissections, we discuss several different dissection problems for the Swiss Cross:

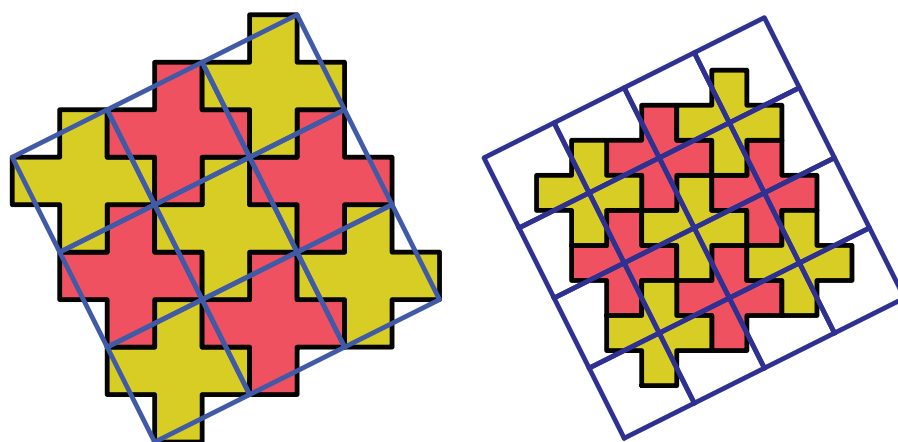


Figure 3.35 Two dissections of the Swiss Cross into a Square

Figure 3.35 shows two dissections of the Swiss Cross into four pieces, which can be assembled into a single square.

Example 3.5.5

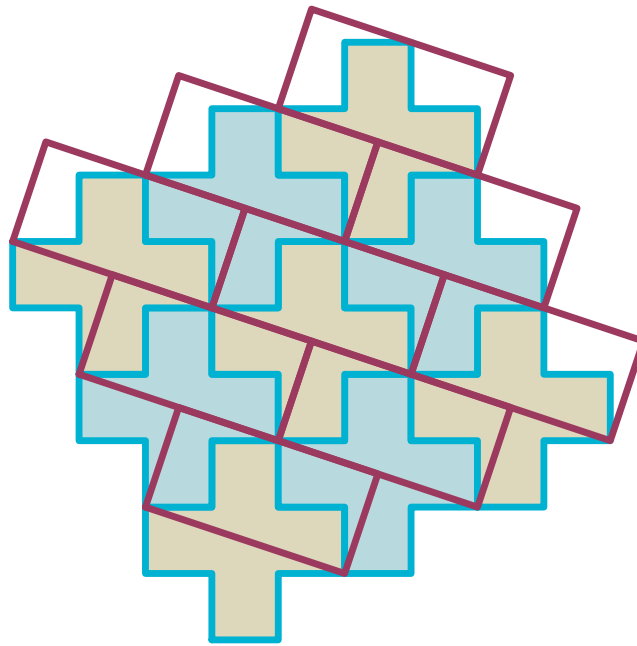


Figure 3.36 Dissections of the Swiss Cross into a Rectangle

It is slightly more difficult to dissect the Swiss Cross into a rectangle that is twice as long as it is tall. This is done with just three tiles in **Figure 3.36**.

Example 3.5.6

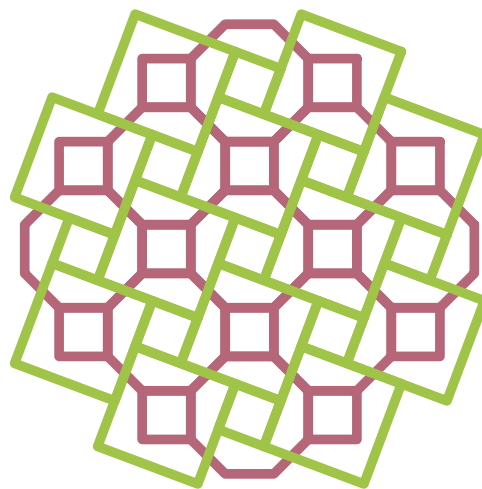


Figure 3.37 Octagon to Square

Figure 3.37 shows a more complicated example that shows how to dissect an octagon into a square. As the octagon itself does not tile the plane, we use instead the Archimedean tiling by octagons and squares together.

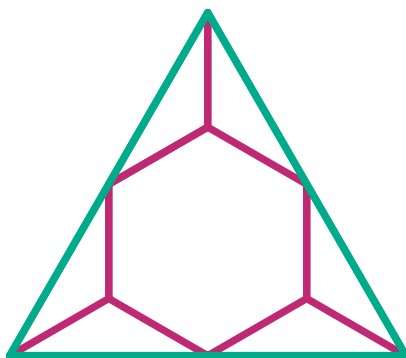


Figure 3.38 Triangle to hexagons

Exercise 3.5.9 In Figure 3.38, the regular tilings by triangles and hexagons were overlaid to find a dissection of a triangle into seven pieces that can be reassembled into two regular hexagons of equal size. Using the same tilings, find a different overlay that shows a dissection into two hexagons with fewer tiles.

Solution: Figure 3.39 shows a six piece solution.

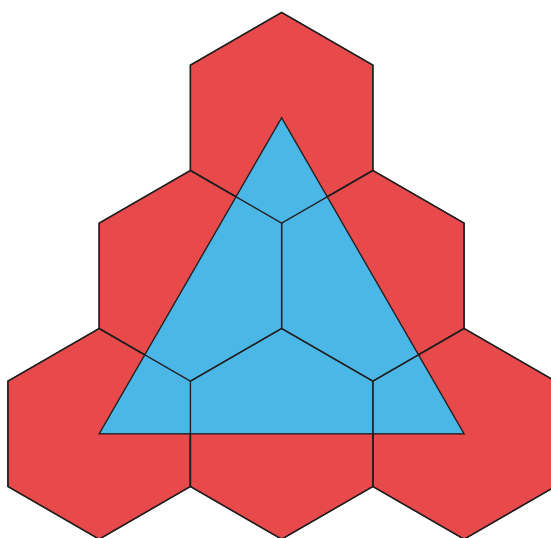


Figure 3.39 Triangle to two hexagons

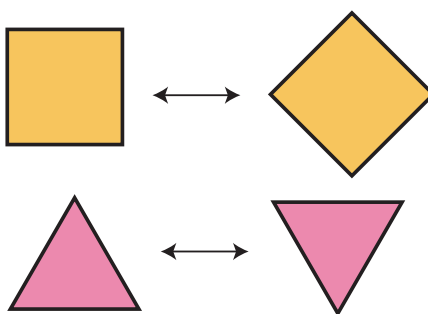


Figure 3.40 Translational Dissections?

Exercise 3.5.10 Is it possible to dissect a square into pieces which can be translated (but not rotated) as to give the same square, albeit rotated by 45° (see [Figure 3.40](#))? Either construct such a translational dissection or show its impossibility.

Solution: The following six-piece solution is due to DOV RHODES:

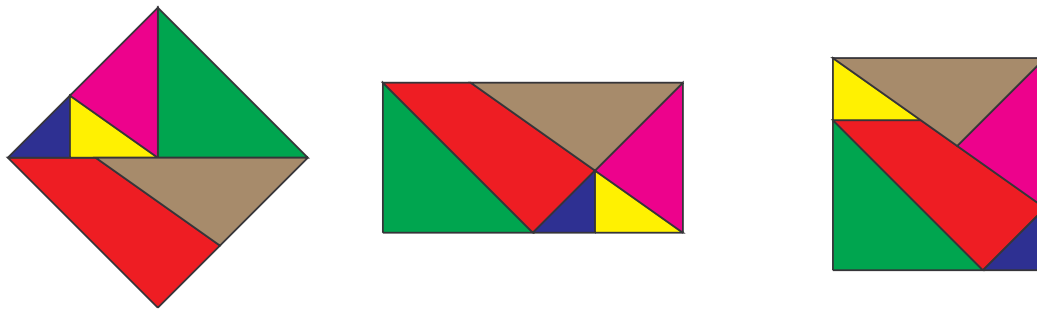


Figure 3.41 Solution to the translational dissection puzzle

Exercise 3.5.11 Is it possible to dissect an equilateral triangle into pieces which can be translated (but not rotated) as to give the same triangle, albeit rotated by 180° (see [Figure 3.40](#))? Either construct such a translational dissection or show its impossibility.

Solution: This is impossible. Any translational dissection of a polygon must preserve (besides the area) another set of invariants: For a polygon, we give each of its edges an orientation by following the polygon boundary counter-clockwise. Then we define the oriented length of a horizontal edge to be its positive length if the edge is oriented from left to right, and otherwise its negative length. Finally, the *horizontal length* of a polygon is the sum of all oriented lengths of its horizontal edges. In particular, our two triangles have horizontal lengths with opposite signs. If we dissect a polygon, the sum of the horizontal lengths of all pieces must be equal to the horizontal length of the polygon. Because the horizontal length is preserved under translation, the two triangles cannot be translationally dissected.

Exercise 3.5.12 A rectangular sheet of paper of dimensions 3×2 needs to be divided into pieces which can be reassembled into two equal rectangular pieces of the same proportions $3 : 2$. The following figures show how to accomplish this using two periodic tilings. Find the explicit coordinates of the indicated points.

Exercise 3.5.13 [Figure 3.42](#) shows two periodic tilings of the plane by swiss crosses that allow to dissect the larger swiss cross into two congruent, smaller swiss crosses, using eight pieces. Using the same tilings, find a dissection using just five pieces.

Solution:

See Figure ??

Exploration 3.5.7 The following figure shows a way to dissect a square into three squares, using 9 pieces.

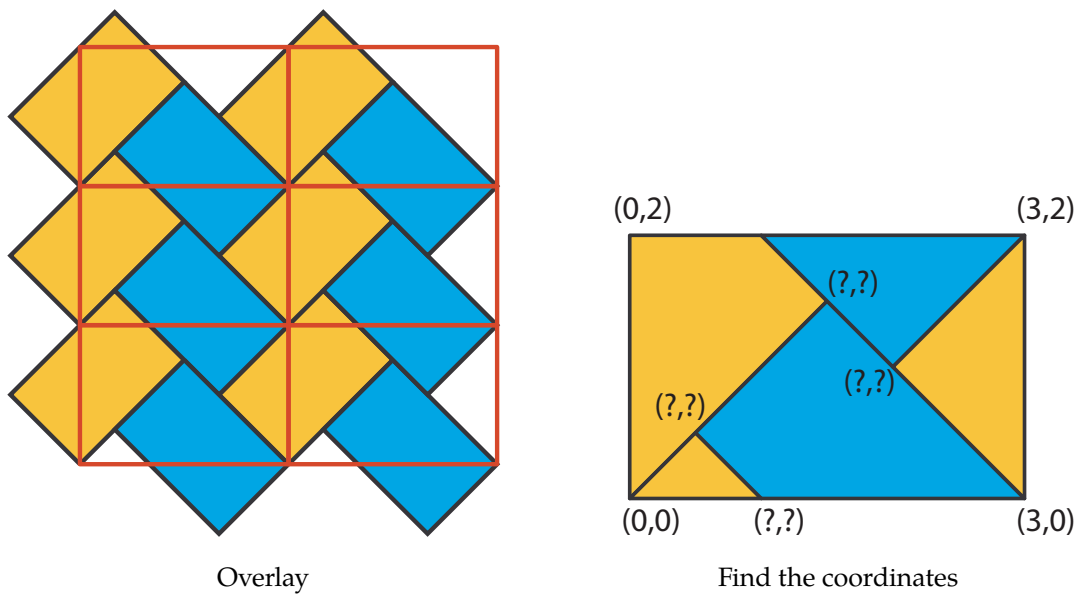


Figure 3.42 Rectangle Doubling

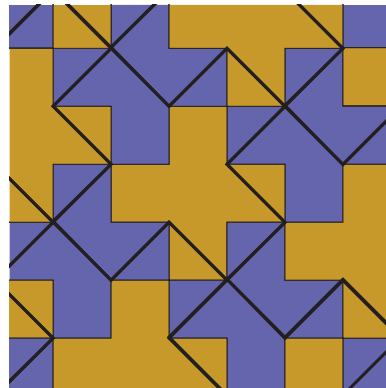


Figure 3.43 Swiss cross dissection with eight pieces

1. Can you find similar dissections using quadrilaterals instead of squares?
2. Can you find a better dissection of a square into 3 squares of the same size? What about into n squares of the same size?
3. Can you find similar dissections of triangles into triangles?
4. What other shapes can you use?

Solution:

See Figure 3.46.

Exploration 3.5.8 We will explore ways in which one rectangle can be dissected into two rectangles.

The easiest case is to start with a rectangle of dimensions $a \times b$, and, using one vertical cut, split it into two rectangles of dimensions $a_1 \times b$ and $a_2 \times b$, where $a_1 + a_2 = a$.

If the two smaller rectangles don't fit together that nicely, one has to try other things. Here are some suggestions:

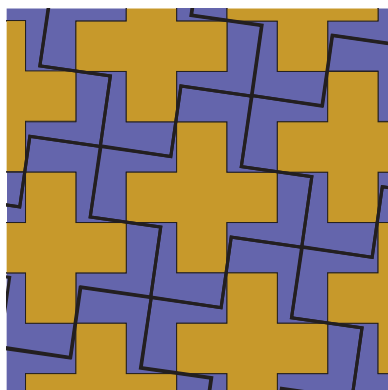


Figure 3.44 Swiss cross dissection with five pieces

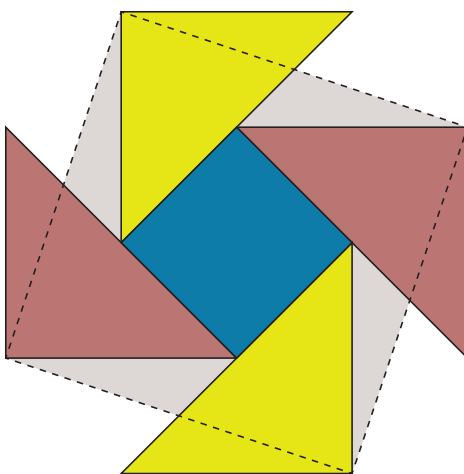


Figure 3.45 Single square dissected into three squares

1. A single rectangle of dimensions 3×2 needs to be divided into pieces which can be reassembled into two equal rectangles of the same proportions $3 : 2$. Find a dissection using no more than five pieces to accomplish this.
2. Can you generalize this to divide any rectangle of dimensions $a \times b$ into two smaller, congruent rectangles of the same proportions $a : b$?
3. How can two rectangles tile the plane periodically?
4. How can just one rectangle tile the plane periodically? Note that the tilings don't have to be edge-to-edge.
5. If the previous problems seem too general, specialize by fixing the proportions of the rectangles to (say) $2 : 1$ or $3 : 2$.
6. Maybe you can find other ways to accomplish such dissections, without relying on the cookie-cutter method?

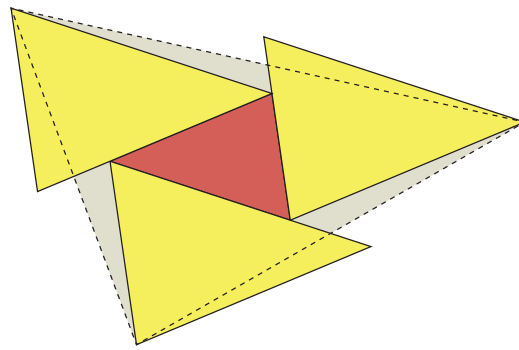


Figure 3.46 Single triangle dissected into four triangles

3.6 POLYFORM DISSECTIONS

In this section, we will apply the overlay method to find numerous dissections of polyforms.

The following exercise collects a few polyomino dissections into squares. To use the tiling method, we need the polyominoes to tile the plane periodically with respect to a square lattice, and the fundamental square must have the same area as the polyominoes. As polyominoes consist of unit squares, the lattice vectors must have integers as entries, so that the area of the polyominoes must be a sum of the squares.

To be more concrete: To dissect a pentomino into a square, the square must have an area of 5. Hence, the edge vectors of the fundamental squares of the square tiling must have length 5 and integer entries. As only $5 = 1^2 + 4^2$, they must be among the eight vectors of the form $(\pm 1, \pm 4)$ and $(\pm 4, \pm 1)$.

Exercise 3.6.1 Find dissections of the following pentominoes into squares with at most the given number of pieces.

1. Find a dissection of a square into the P-pentomino with at most four pieces.
2. Find a dissection of a square into the Y-pentomino with at most four pieces.
3. Find a dissection of a square into the Z-pentomino with at most three pieces.

Solution: Tile the plane with the pentomino periodically as below. The shown squares tile the plane with the same period lattice, and overlaying the two tilings gives a dissection as desired.

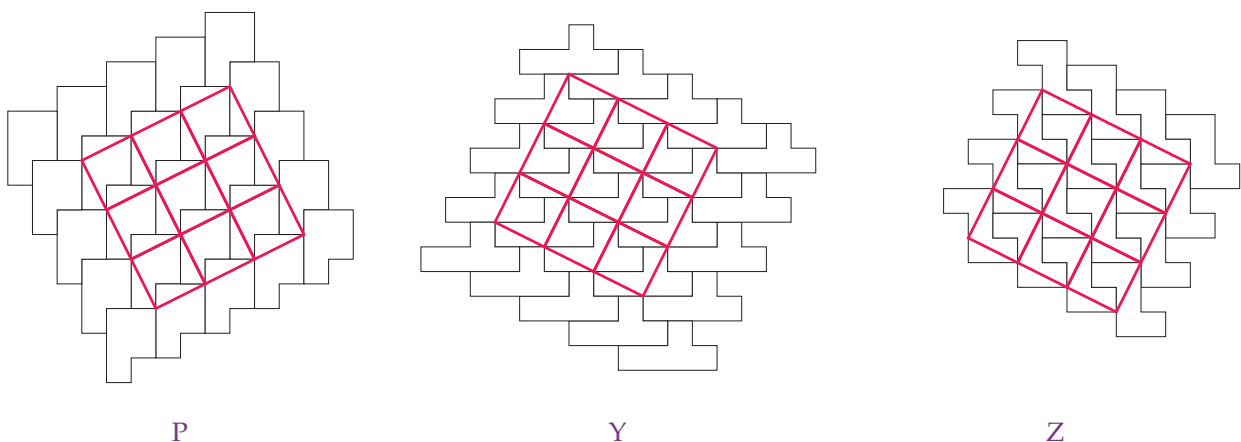


Figure 3.47 Some possible tilings with pentominoes

Exercise 3.6.2 Find dissections of a square into two equal pentominoes using at most the given number of pieces. Note that you will need to find a tiling by the pair of pentominoes.

1. Find a dissection of a square into two R-pentominoes, using at most five pieces.
2. Find a dissection of a square into two Z-pentominoes, using at most four pieces.
3. Find a dissection of a square into two Y-pentominoes, using at most five pieces.
4. Find a dissection of a square into two U-pentominoes, using at most five pieces.
5. Find a dissection of a square into two P-pentominoes, using at most five pieces.
6. Find a dissection of a square into two I-pentominoes, using at most six pieces.

Solution: Tile the plane with the pentominoes as in the figures. The shown squares tile the plane with the same period lattice, and overlaying the two tilings gives a dissection as desired.

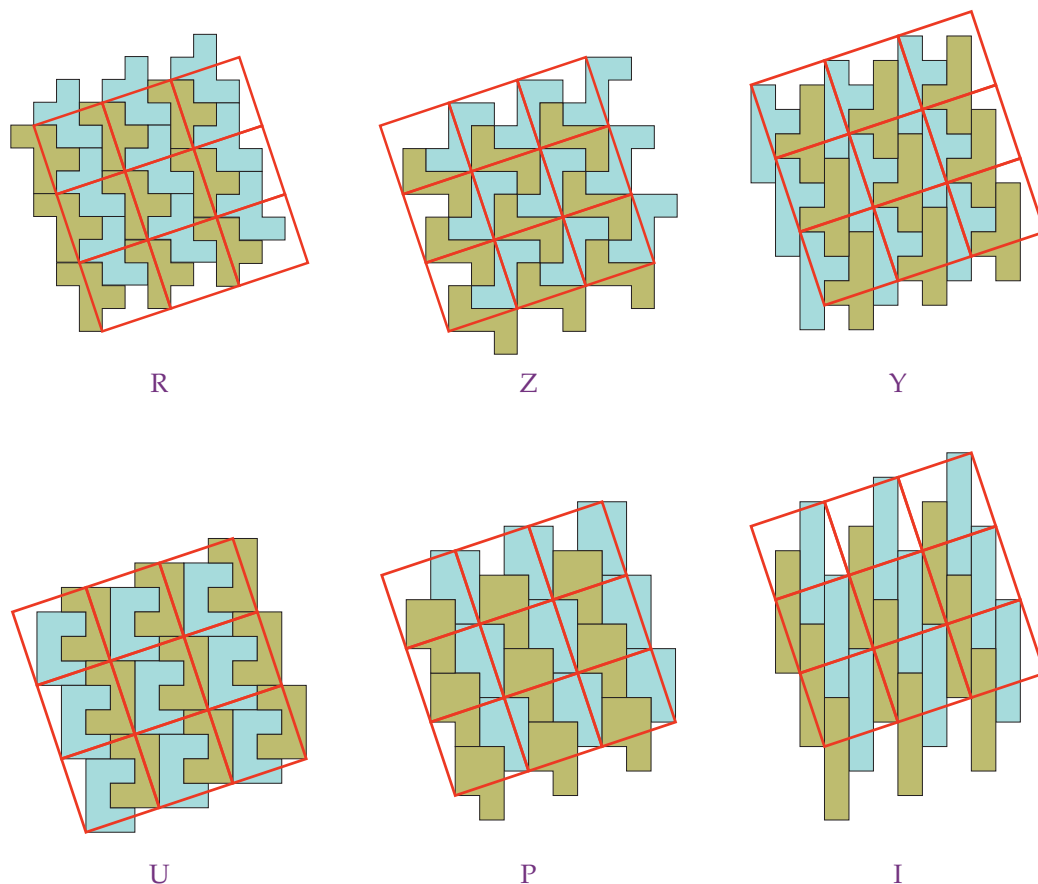


Figure 3.48 Square into two pentominoes

Exercise 3.6.3 Find dissections of a square into two different pentominoes using at most the given number of pieces. Note that you will need to find a tiling by the pair of pentominoes.

1. Find a dissection of a square into a L-pentomino and a P-pentomino, using at most four pieces.
2. Find a dissection of a square into a P-pentomino and a T-pentomino, using at most four pieces.
3. Find a dissection of a square into a R-pentomino and a U-pentomino, using at most five pieces.

Solution: Tile the plane with the pentominoes as in the figures. The shown squares tile the plane with the same period lattice, and overlaying the two tilings gives a dissection as desired.

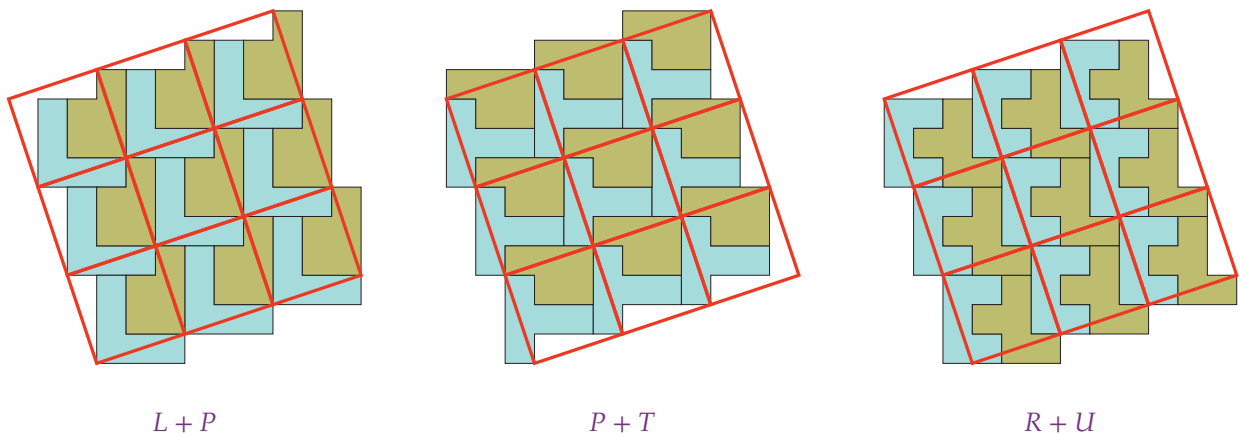


Figure 3.49 Square into two different pentominoes

Exercise 3.6.4 Show that a single hexomino cannot tile the plane periodically just by translations so that the lattice is the square lattice.

Solution: If this were possible, the fundamental square of the square lattice would have to have area 6, and its edge vectors had integer entries. But there is no solution of $a^2 + b^2 = 6$ in integers.

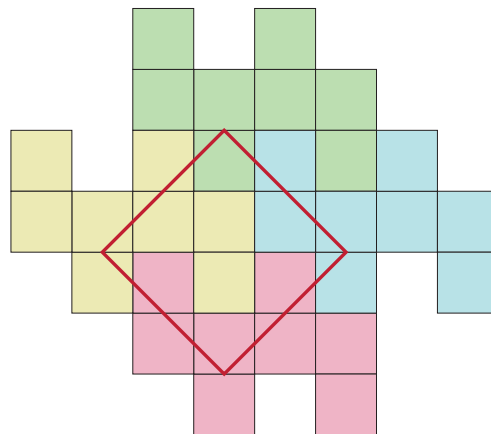


Figure 3.50

Exercise 3.6.5 Find three different octominoes, each of which tiles the plane periodically by just translating that octomino, and has as a lattice the square lattice. An example is given in Figure 3.50. Use this to find a dissection of the square of edge length $\sqrt{8}$ into your octominoes.

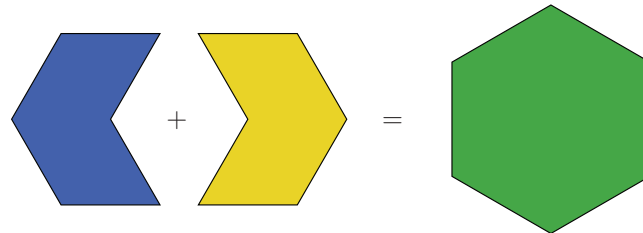


Figure 3.51

Exercise 3.6.6 Dissect a regular hexagon into two V-tetramonds, using at most five pieces. The direction of the tetramonds does not matter.

Solution: See Figure 3.52.

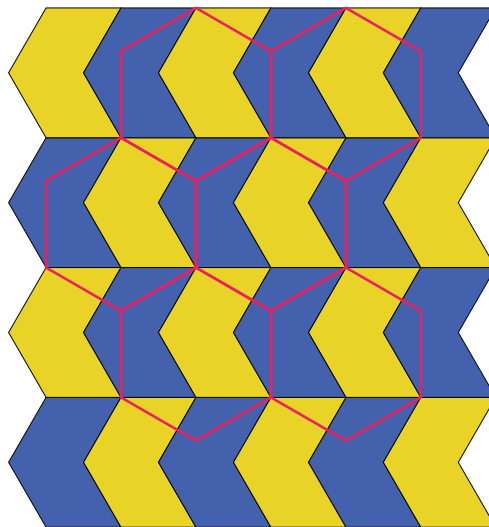


Figure 3.52

Exercise 3.6.7 Show that there cannot be a periodic tiling by heptiamonds that just translates a single heptiamond (i.e., a shape with seven equilateral triangles, glued together edge to edge).

Exploration 3.6.1 There are twelve Hexiamonds, see Figure 2.19.

One goal of this exploration is to find dissections of a hexagon into three hexiamonds with as few pieces as possible. These can be equal or different hexiamonds.

A possible strategy is to pick three hexiamonds and to assemble them into a single tile that tiles the plane periodically with a hexagonal lattice. Then the overlay argument will create dissections.

Another goal is to prove that certain combinations of three hexiamonds cannot tile the plane periodically. For instance, you might want to prove that no combination of three *P*-hexiamonds will tile the plane periodically.

Exploration 3.6.2 In this exploration, you will discover dissections of squares into a pentomino and a tromino. These two tiles together have an area of 8 unit squares. To use the dissection method, you will therefore need to find a tiling by the pentomino and the tromino that is periodic with respect to the lattice generated by $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ and $\begin{pmatrix} 2 \\ -2 \end{pmatrix}$. To get started, pick your favorite pentomino, and use the translations from the lattice to create a partial tiling of the plane. There will be gaps left, which you need to fill with a tromino. This might be impossible or possible in several ways. Then, by shifting the corresponding tiling by squares of area 8, find a dissection of the square into your pentomino and tromino.

3.7 NAPOLEON'S THEOREM

The goal of this section is to prove the following

Theorem 3.7.1 *Let ABC be a triangle. Construct equilateral triangles CBA' , ACB' and BAC' outside ABC . Then the centers O_A , O_B , O_C of the new triangles form another equilateral triangle.*

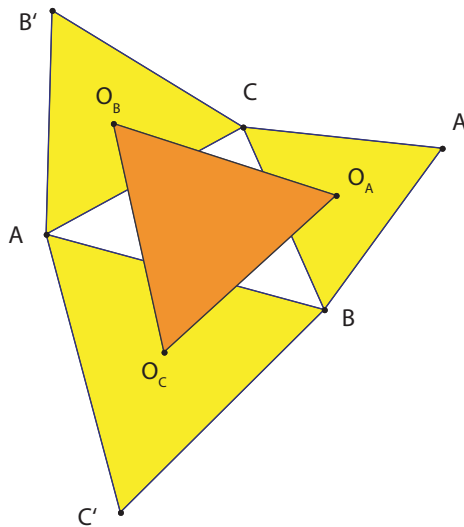


Figure 3.53 Napoleon's Theorem

It is not known whether Napoleon actually proved this theorem. He was interested in Mathematics and discussed science with Lagrange and Laplace. The first known written record is an article by W. Rutherford in 1825, which doesn't mention Napoleon.

We will prove this result in two different ways.

The first approach is computational and uses complex numbers, while the second will use tilings. We begin with a simple characterization of equilateral triangles, using complex numbers:

Proposition 3.7.2 *A triangle ABC is equilateral if and only if*

$$A + \zeta B + \zeta^2 C = 0$$

where

$$\zeta = \text{cis}(120^\circ) = -\frac{1}{2} + \frac{\sqrt{3}}{2}i$$

Proof. Let's first assume that ABC is equilateral, and assume further that the center of ABC is the origin. Then $B = \zeta A$ and $C = \zeta^2 A$ so that

$$\begin{aligned} A + \zeta B + \zeta^2 C &= (1 + \zeta + \zeta^2) A \\ &= 0 \end{aligned}$$

so that the claim holds in this special case. For a general equilateral triangle ABC , let O be its center and define a new triangle by

$$A' = A - O \quad B' = B - O \quad C' = C - O$$

Then $A'B'C'$ has the origin as its center, so

$$\begin{aligned} 0 &= A' + \zeta B' + \zeta^2 C' \\ &= A + \zeta B + \zeta^2 C - (1 + \zeta + \zeta^2) O \\ &= A + \zeta B + \zeta^2 C, \end{aligned}$$

so that the equation also holds in general.

Now assume that ABC is a triangle for which $A + \zeta B + \zeta^2 C = 0$. If T is any complex number and we define

$$A' = A - T \quad B' = B - T \quad C' = C - T.$$

Then, by the same computation as above, we also have $A' + \zeta B' + \zeta^2 C' = 0$. Let us take, for instance, $T = -C$. Then it follows that $A' + \zeta B' = 0$. This means that B' can be obtained from A' by a 60° rotation around C' . In particular, $\angle A'C'B' = 60^\circ$. As ABC is congruent to $A'B'C'$ by translation, we also have that $\angle ACB = 60^\circ$. A similar argument shows that the other angles are 60° as well, and ABC is equilateral. $\square$

Now we can prove Napoleon's theorem.

Proof. Observe that the centers of the Napoleon triangles can be written as

$$\begin{aligned} Z_A &= \frac{1}{3}(C + B + O_A) \\ Z_B &= \frac{1}{3}(A + C + O_B) \\ Z_C &= \frac{1}{3}(B + A + O_C) \end{aligned}$$

We now use our criterion and compute

$$\begin{aligned} 3(Z_A + \zeta Z_B + \zeta^2 Z_C) &= (C + B + O_A) + \zeta(A + C + O_B) \\ &\quad + \zeta^2(B + A + O_C) \\ &= (O_A + \zeta C + \zeta^2 B) + (C + \zeta O_B + \zeta^2 A) + (B + \zeta A + \zeta^2 O_C) \\ &= 0 + 0 + 0 \\ &= 0 \end{aligned}$$

There is another, quite different proof of Napoleon's theorem, using a tessellation argument:

Proof. The key observation is that one can extend the Napoleon figure to a tiling of the Euclidean plane, consisting only of copies of the original triangle and the three equilateral triangles. Here, each equilateral triangle is surrounded by copies of the original triangle in a rotationally symmetric way. As illustrated [Figure 3.54](#), this implies that the centers of equilateral triangles of one color (say blue) are the vertices of a hexagonal tiling by equilateral triangles. Moreover, the centers of the triangles of the two other colors (say green and purple) are the midpoints of these triangles.

This implies that the centers of all equilateral triangles together form a finer hexagonal tiling. In particular, the centers of the three Napoleon triangles form an equilateral triangle. $\square$

Exercise 3.7.3 Show that Napoleon's Theorem remains valid if the equilateral triangles are constructed inside ABC .

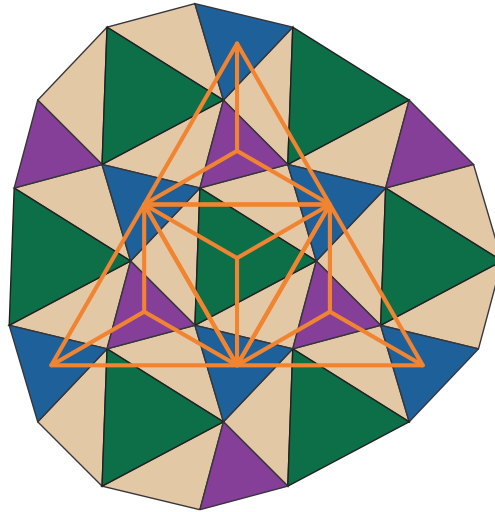


Figure 3.54 Napoleon's Theorem

Exercise 3.7.4 Show the following generalization of Napoleon's Theorem:

Theorem 3.7.5 Let ABC be a triangle. Construct triangles CBA' , ACB' , and BAC' all similar to a fixed triangle XYZ outside of ABC so that the edges of the new triangles common to ABC all correspond to the same edge of XYZ under the similarity. Then the centers O_A , O_B , O_C of the new triangles form another triangle similar to XYZ .

3.8 MORLEY'S THEOREM

We will now discuss a rather surprising and famous theorem that has attracted many famous mathematicians. It was first discovered in 1899 by Frank Morley.

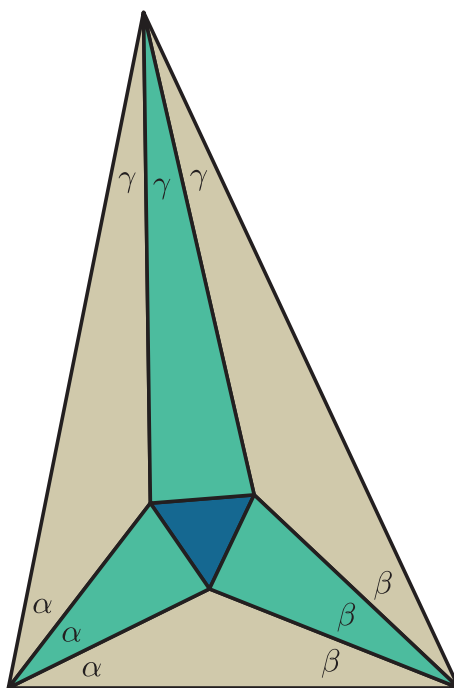


Figure 3.55 Morley's Theorem

Theorem 3.8.1 *The angle trisectors in a Euclidean triangle meet at the corners of an equilateral triangle.*

This should remind us of the very well-known

Theorem 3.8.2 *The angle bisectors in a Euclidean triangle meet in one point.*

This theorem becomes immediately clear when we consider the center of the incircle of the triangle. Points on an angle bisector stay at the same distance from the two legs of the angle, and therefore the three bisectors must meet in a point that has equal distance from all three sides of the triangle — which is the center of the incircle.

The annoying problem with Morley's theorem is that there is no such simple proof for it.

The proof we are going to discuss is due to John H. Conway and makes clever use of a dissection argument. Before we do so, however, we will reprove the bisection theorem using an argument similar to the one in Conway's proof.

Let's introduce the notation $\phi^* = \phi + 90$ for this argument. Suppose we are given a triangle with angles $2\alpha, 2\beta, 2\gamma$. Let's construct three triangles. The first has angles α, β, γ^* , call the vertices A, B, P . Draw the perpendicular from P to AB and make it of length 1. Similarly, construct triangle with angles α, β^*, γ and α^*, β, γ and make their heights to be of length 1 as well. The three triangles fit together around the point P , because adjacent triangles are cut by their heights into smaller triangles so that the two adjacent smaller triangles are congruent by reflection at their common edge.

This proof is not too different from the first argument we gave, but it is actually better, because it also demonstrates the existence of the incircle, something that we had assumed silently.

Let A, B, C be the vertices of the given triangle, with angles $3\alpha, 3\beta, 3\gamma$. For convenience, let's introduce the notation

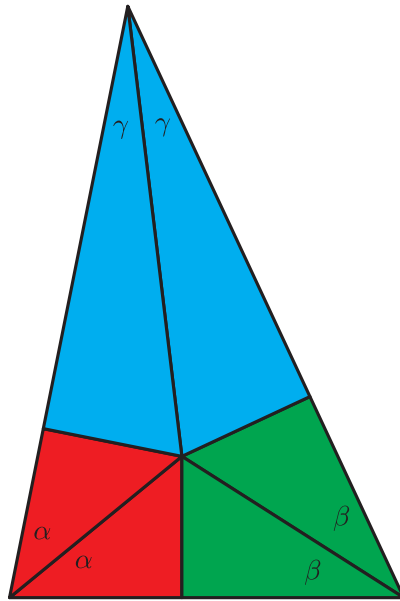


Figure 3.56 Angle bisectors meet at center of incircle

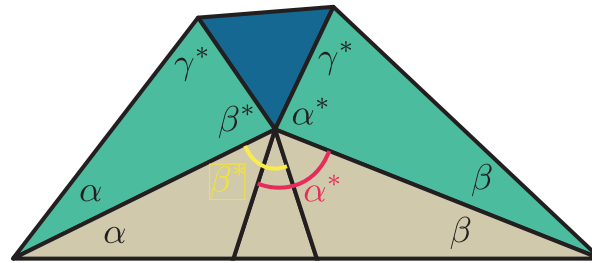


Figure 3.57 Angle bisectors meet at the center of the incircle

$$\phi^* = \phi + 60^\circ$$

For example, we have

$$\alpha + \beta + \gamma = 0^*$$

We will now construct seven triangles that will eventually become the seven triangles in the Morley triangle. We will specify all angles in the triangle, and then their sizes.

1. $0^*, 0^*, 0^*$ is an equilateral triangle of edge length 1.
2. $\alpha, \beta^*, \gamma^*$. Note that the angles add up to 180° , so the triangle exists. We determine its size by specifying that the edge opposite to angle α has length 1.
3. $\alpha^*, \beta, \gamma^*$ and

4. $\alpha^*, \beta^*, \gamma$ are defined correspondingly. These last three triangles fit together with the first (equilateral) triangle so that together they look like the green and blue triangles in Morley's configuration. Observe that we have yet no clue that the angles we used to define the green triangles are the correct angles for the Morley configuration.
5. α, β, γ^* . To normalize its size, call the vertices of the triangle A, B, P . Draw the two lines from P that intersect AB at the angle γ^* (note that $\alpha + \gamma^*$ and $\beta + \gamma^*$ are both $< 180^\circ$ so that we can really draw these lines). Make the two segments from P to the intersections with AB of length 1. The triangle ABP is then cut into an isosceles triangle and two triangles which are congruent by reflection to the triangles $\alpha, \beta^*, \gamma^*$ and $\alpha^*, \beta, \gamma^*$. In particular, the triangle α, β, γ^* fits right in between these two.
6. α, β^*, γ and
7. α^*, β, γ are again defined the same way and also fit as khaki triangles into our puzzle.

4 Playing with Graphs and Maps

A *graph* is a fundamental mathematical construct that tracks whether two objects are *connected* or not in some sense. Graphs arise in many different areas and have many applications. In the context of games, they find twofold use: As game boards, and as the graph of all possible positions of a given game, where two positions are connected if there is a move from one to the other.

4.1 GRAPHS

In this section, we introduce the very fundamental concept of a *graph*, give examples, and study basic properties.

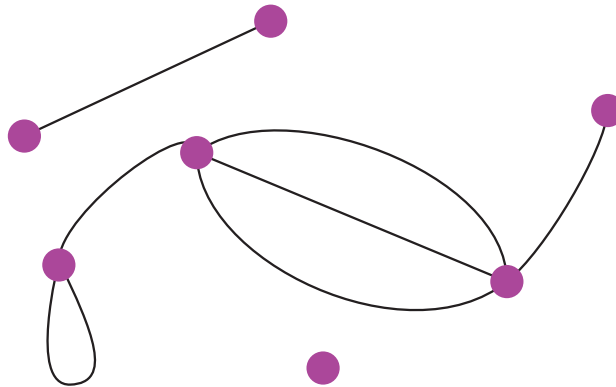


Figure 4.1 A graph with multiple edges and a loop

A graph is intended to represent a set of objects that are either mutually related or unrelated. For instance, the objects could be web pages that are related if there is a link from one to the other. In a graph, the objects are represented as points called *vertices*, and the relations by *edges*. Formally, this is done as follows:

graph

Definition 4.1.1 A graph consists of a finite set V of vertices and a set of edges E together with two maps $f_1, f_2 : E \rightarrow V$. The edge $e \in E$ is said to connect the vertices $f_1(e)$ and $f_2(e)$, and these vertices are then called adjacent.

Thus, a graph can have multiple edges connecting the same vertices, loops that connect a vertex with itself, and isolated vertices that are not connected to any other point. Sometimes it is useful to consider also *directed graphs* where the order of the *endpoints* $f_1(e), f_2(e)$ of an edge is used to impose a direction on the edges. Then $f_1(e)$ is the *predecessor* of $f_2(e)$, while $f_2(e)$ is the *successor* of $f_1(e)$.

Most often, we will be happy with the slightly narrower concept of a *simple graph* that avoids multiple edges and loops:

simple
graph

Definition 4.1.2 A simple graph consists of a finite set V of vertices and a set of edges E which is a set of unordered pairs of vertices. Two vertices v_1 and v_2 are called adjacent if the pair (v_1, v_2) belongs to the set of edges, and that edge is said to connect the vertices v_1 and v_2 . Usually, simple graphs do not allow loops, i.e., edges of the form (v_1, v_1) .

For instance, we could define the simple graph shown in **Figure 4.2** by a vertex set $V = \{1, 2, 3, 4, 5, 6\}$ and an edge set

$$E = \{\{1, 2\}, \{1, 4\}, \{1, 6\}, \{2, 3\}, \{2, 5\}, \{3, 4\}, \{3, 6\}, \{4, 5\}, \{5, 6\}\}.$$

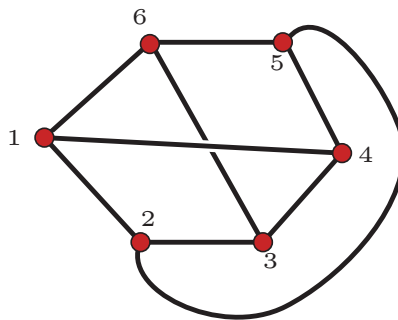


Figure 4.2 A simple graph Γ

A typical way to describe a graph is by a *graph drawing*, see figures 4.1 and 4.2. Sometimes it is helpful to label the vertices in order to be able to refer to them.

Note that the intersection of the edges 14 and 36 is not a vertex of the graph in the example to the right.

Another way to describe a simple graph is by its *adjacency matrix*:

adjacency
matrix

Definition 4.1.3 Let Γ be a simple graph with vertices $V = \{v_1, \dots, v_n\}$. Define a $n \times n$ matrix $A = (a_{i,j})$ by setting

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ and } v_j \text{ are adjacent,} \\ 0 & \text{otherwise} \end{cases}$$

For example, the simple graph in **Figure 4.2** has the adjacency matrix

$$\begin{pmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{pmatrix}$$

According to the definition of a simple graph, its adjacency matrix has only zeroes or ones as entries, is symmetric, and has zeroes on the diagonal. On the other hand, given any symmetric $n \times n$ matrix A with these properties, there is a simple graph with that adjacency matrix: We just define the set of vertices to be the numbers $\{1, \dots, n\}$ and the set of edges to be those pairs $\{i, j\}$ where $a_{ij} = 1$.

Sometimes one allows for a vertex to be connected to itself by a simple loop; this leads to a diagonal entry 1 for that vertex in the adjacency matrix.

We now give some important examples:

complete
graph

1. The *complete graph* K_n has as vertices the set $\{1, \dots, n\}$, and all vertices are connected to each other by an edge.
2. For $n \geq 2$, the *path graph* P_n has as vertices the set $\{1, \dots, n\}$, and two vertices are connected if and only if $|i - j| = 1$.
3. For $n \geq 3$, the *simple circuit* C_n has as vertices the set $\{1, \dots, n\}$, and two vertices are connected if and only if $|i - j| = 1$ or $|i - j| = n - 1$. The complete graph K_3 is also the circuit C_3 .

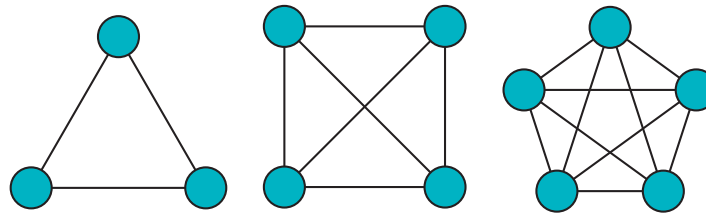


Figure 4.3 The complete graphs K_3 , K_4 , and K_5



Figure 4.4 The path graph P_8

4. Take two disjoint sets A and B with m and n elements, respectively. The vertices of the *complete bipartite graph* $K_{m,n}$ is the union $A \cup B$. All pairs $\{a, b\}$ with $a \in A$ and $b \in B$ are edges. In other words, no two vertices from the same subset are joined, but all pairs from different subsets.

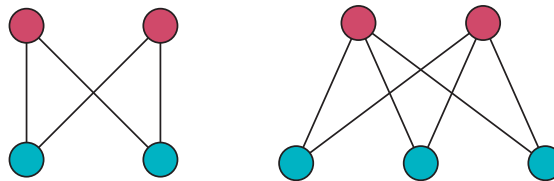


Figure 4.5 The complete bipartite graphs $K_{2,2}$ and $K_{2,3}$

- star 5. The complete bipartite graphs $K_{1,n}$ are also called *stars* and denoted by S_n .

bipartite More generally, we define *bipartite* graphs as follows:

Definition 4.1.4 A graph is called *bipartite* if the set of vertices is a disjoint union of sets A and B and every edge connects a vertex from A with a vertex from B .

Exercise 4.1.5 How many edges do the complete graph K_n and the complete bipartite graph $K_{m,n}$ have?

Solution: The complete graph K_n has $\frac{1}{2}n(n-1)$ edges, because each of the n vertices is connected with $n-1$ other vertices. This gives $n(n-1)$ edges, but we are counting each edge twice.
The complete graph $K_{m,n}$ has $m \cdot n$ edges. Here, each of the m vertices is connected to each of the n vertices.

Exercise 4.1.6 Draw an example of a simple graph with eight vertices such that there are two vertices of degree 1, 2, 3, and 4, respectively.

Solution: See Figure 4.6

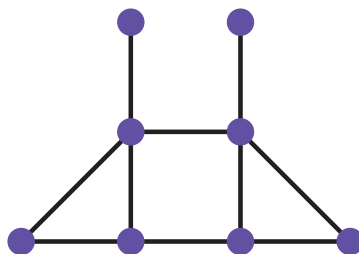


Figure 4.6

Definition 4.1.7 A path of length $n - 1$ is a sequence of vertices $v_1 v_2 \dots v_n$ such that consecutive vertices are adjacent. A graph is connected if there is a path from any vertex to any other vertex.

Given a graph Γ with adjacency matrix A , the matrix product $A \cdot A$, and more generally, the matrix powers A^k allow us to count how many different paths there are between different vertices.

To see this, let $B = A^2$, and let b_{rs} be the entry of B in row r and column s . Assume also that Γ has n vertices, so that both A and B are $n \times n$ matrices. Recall that by the definition of the matrix product,

$$b_{rs} = a_{r1}a_{1s} + a_{r2}a_{2s} + \dots + a_{rn}a_{ns}$$

A general summand $a_{ri}a_{is}$ is equal to 1 if and only if there is a connection from vertex r to vertex s via vertex i , and 0 otherwise. Thus, the sum counts how many different 2-step connections there are from vertex r to vertex s .

More generally, we have (for the same reason)

Proposition 4.1.8 Let Γ be a graph with adjacency matrix A . The entry at position (r, s) of the matrix power A^k counts the number of different paths of length k from vertex r to vertex s .

Exercise 4.1.9 Determine for the complete graph K_3 how many different 5-step connections there are from a given vertex back to itself.

Solution: The complete graph K_3 has as its adjacency matrix

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Its fifth power is

$$A^5 = \begin{pmatrix} 10 & 11 & 11 \\ 11 & 10 & 11 \\ 11 & 11 & 10 \end{pmatrix},$$

so that there are 10 connections from a fixed vertex to itself.

There are many ways to attach interesting numerical quantities to a graph. The simplest one is the *valency* of a vertex:

path
connected

valency
degree

Definition 4.1.10 The valency or degree of a vertex in a graph is the number of adjacent vertices.

Exercise 4.1.11 Find as many graphs as you can with 6 vertices where each vertex has degree 2. Which of your graphs are simple?

Solution: The only simple solutions are the circuit C_6 and the union of two circuits $\{C_3, C_3\}$. In general, a solution decomposes into a union of circuits whose lengths add up to 6. Thus each solution corresponds to a *partition* of the integer 6 in a sum of smaller integers. There are 11 such partitions:

$$\begin{aligned} 6 &= 6 \\ &= 1 + 5 \\ &= 2 + 4 \\ &= 3 + 3 \\ &= 1 + 1 + 4 \\ &= 1 + 2 + 3 \\ &= 2 + 2 + 2 \\ &= 1 + 1 + 1 + 3 \\ &= 1 + 1 + 2 + 2 \\ &= 1 + 1 + 1 + 1 + 2 \\ &= 1 + 1 + 1 + 1 + 1 + 1 \end{aligned}$$

Exercise 4.1.12 Find the two connected simple graphs with 6 vertices where each vertex has degree 3.

Solution:

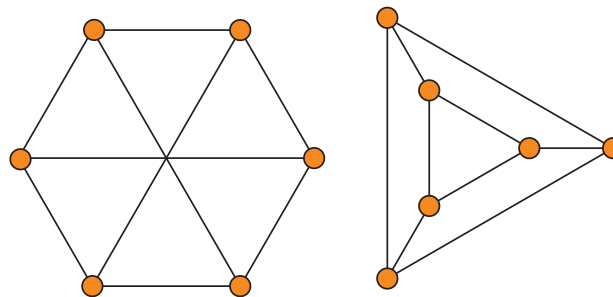


Figure 4.7 The two cubic graphs with six vertices

Exercise 4.1.13 Can there be a graph with 11 vertices such that every vertex has degree 3?

Solution: No, this would contradict the **Handshaking Lemma 4.1.14**. The number of edges of such a graph would have to be $33/2$.

There is a simple relationship between the degrees of the vertices and the number of edges of a graph:

Lemma 4.1.14 (Handshaking Lemma) *Let Γ be a graph with v vertices of degrees $d_1, \dots, d_v$ and e edges. Then*

$$2e = d_1 + \dots + d_v$$

Proof. When we add the degrees of all vertices, we count all edges twice. □

Corollary 4.1.15 *If all vertices in a graph have at least valency 3, then $2e \geq 3v$.*

Exercise 4.1.16 *Suppose a graph has 15 edges, 3 vertices of degree 5, and all other vertices of degree 3. How many vertices does the graph have?*

Solution: Let n be the number of vertices of degree 3. By the handshaking lemma, $2 \cdot 15 = 5 \cdot 3 + 3 \cdot n$ so that $n = 5$. Thus, there are a total of 8 vertices.

Exercise 4.1.17 *Show that a simple graph with $v > 1$ vertices has at least two vertices of the same degree.*

Solution: Suppose there is a simple graph with $v > 1$ where all vertices have different degrees. As the graph is simple, the only possible degrees are the numbers $0, 1, 2, \dots, v - 1$. This means we have as many available degrees as there are vertices, and because all vertices have different degrees, all available degrees must occur. Consider the vertex of degree $v - 1$. It is connected to all other vertices, in particular the one with degree 0, which is not possible. This shows that a simple graph with these properties can't exist.

Exercise 4.1.18

1. *At a party, there are 10 people. One has shaken one hand, two have shaken two hands, three have shaken three hands, and yes, you guessed it, four have shaken four hands each. No two people shook hands more than once. Can this be? Explain!*
2. *At a party, there are 15 people. One has shaken one hand, two have shaken two hands, etc., and the last five have shaken five hands each. No two people shook hands more than once. Can this be? Explain!*

Solution: See Figure 4.8.

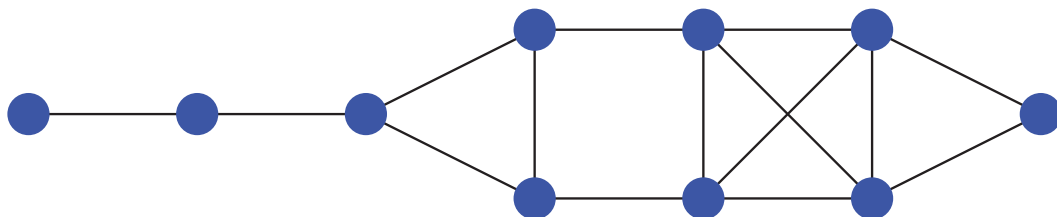


Figure 4.8 Handshake solution

1. Yes, this is possible, see Figure 4.8.
2. By the handshaking lemma, $\sum_{i=1}^5 i^2 = 2e$ so that $2e = 55$, which is impossible.

For bipartite graphs, there is a stronger form of the handshaking lemma:

Lemma 4.1.19 *In a bipartite graph with vertex set $A \cup B$ where every edge connects a vertex from A to a vertex from B ,*

$$\sum_{a \in A} d_a = \sum_{b \in B} d_b .$$

Proof. Both sides of the equation count the number of edges of the graph. □

Exercise 4.1.20 *At a party, there are eight people. Four of them belong to a group called Even, the other four to a group called Odd. They are shaking hands according to the following rules: Every person shakes at least one hand, nobody shakes his or her own hand, or that of anybody from his or her own group. No two people shake hands twice. Furthermore, every odd person shakes an odd number of hands, while every even person shakes an even number of hands. We also know that not all even people shake just two hands, and not all odd people shake just three hands. How many handshakes were there in total?*

Solution: Consider the graph which has as vertices the eight people, and edges indicate handshakes. This graph is bipartite by assumption. The valencies of the odd people are either 1 or 3, and the valencies of the even people are either 2 or 4. The possible sums of the odd valencies are $1 + 1 + 1 + 1 = 4$, $1 + 1 + 1 + 3 = 6$, $1 + 1 + 3 + 3 = 8$, $1 + 3 + 3 + 3 = 10$, and $3 + 3 + 3 + 3 = 12$. Similarly, the possible sums of even valencies are $2 + 2 + 2 + 2 = 8$, $2 + 2 + 2 + 4 = 10$, $2 + 2 + 4 + 4 = 12$, $2 + 4 + 4 + 4 = 14$, and $4 + 4 + 4 + 4 = 16$. These sums are the number of edges and thus must match. The only matching pairs are

$$1 + 1 + 3 + 3 = 2 + 2 + 2 + 2$$

$$1 + 3 + 3 + 3 = 2 + 2 + 2 + 4$$

$$3 + 3 + 3 + 3 = 2 + 2 + 4 + 4$$

The first and last are excluded by the conditions, hence the middle one is the solution. There are 10 handshakes.

Exercise 4.1.21 *At a party, there are seven people. Group A consists of three people, and the other four make up group B. After shaking hands, they claim the following: Nobody from group A has shaken hands with anybody from group A, and nobody from group B has shaken hands with anybody from group B. Moreover, everybody from group A has shaken three or four hands, while everybody from group B has shaken one or two hands. Can this be?*

Solution: Consider the graph that has as its vertices the eight people, and the edges indicate handshakes. This graph is bipartite. The valencies of the A people are either 3 or 4, and the valencies of the B people are either 1 or 2. As the graph is bipartite, the sum of the valencies on either side gives the number of edges. However, that sum is maximally 8 for the B -side and minimally 9 for the A -side, a contradiction.

The next rather famous problem goes reportedly back to PAUL RICHARD HALMOS.

Exercise 4.1.22 *Alice and Bert go to a party where there are four other couples. Each guest shakes hands once with everybody they are not acquainted with. Then Alice asks everybody else how often they shook hands. She gets nine different answers. How often did Bert shake hands?*

Solution: Because nobody can shake hands more than eight times (everybody is acquainted with their partner, hopefully), the answers Alice gets can range between 0 and 8. As the answers are all different, the nine answers she gets are precisely the set of numbers from 0 to 8. Let's call each person Zero, One, Two, ..., Eight, depending on their answers. One of them is Bert, and Alice is not among them.

Now we find out who the couples are. Eight must be the partner of Zero, as Eight shook hands with everybody except the one person they are acquainted with. Therefore, all persons they is not acquainted with have shaken hands at least once, and the only remaining person is Zero.

Now let's look at Seven. They did not shake hands with One, because One must have used their only handshake with Eight. Thus, Seven must have shaken hands with everybody except Zero and One, which makes One their partner.

The same argument shows that Six is the partner of Two, and Five is the partner of Three. This leaves Four as Bert, the partner for Alice. Hence, Bert has shaken hands four times.

Exercise 4.1.23 *Alex and Berta go to a party where each person shakes hands with everybody else. Alex shook hands with three times as many men as women, and Berta shook hands with four times as many men as women. How many people were at the party?*

Solution: Let m be the number of men and w be the number of women at the party. Then, according to Alex, $3w = m - 1$, and according to Berta, $m = 4(w - 1)$. Solving for m and w gives $w = 5$ and $m = 16$, so that altogether there were 21 people at the party.

Exercise 4.1.24 *Alex and Berta go to another party where each woman shook hands with each man, but no two women nor two men shook hands with each other. There was a total of 56 handshakes, and nobody shook hands more than 12 times. How many people were at the party?*

Solution: Let m be the number of men and w be the number of women at the party. The handshaking graph is complete bipartite, so there were $m \cdot w = 56$ handshakes. As m and w are both at most 12, this leaves $\{m, w\} = \{7, 8\}$. Hence there were 15 people at the party.

We will consider two graphs as equal (or *isomorphic*), if we can match up the vertices and edges of both graphs. More precisely:

Definition 4.1.25 *Let Γ_1 and Γ_2 be graphs with vertex sets V_1 and V_2 . A bijective map $f : V_1 \rightarrow V_2$ is called a graph isomorphism if for all vertices $v_1, v_2 \in V_1$, v_1 and v_2 are adjacent in Γ_1 if and only if $f(v_1)$ and $f(v_2)$ are adjacent in Γ_2 .*

It can be quite difficult to determine whether two graphs are isomorphic. **Figure 4.9** shows a simple example — two different drawings of the edge graph of a cube.

Exercise 4.1.26 *Draw all (non-isomorphic) connected simple graphs with four vertices.*

Solution:

Informally, a *subgraph* is obtained by removing edges and vertices from a graph. More precisely:

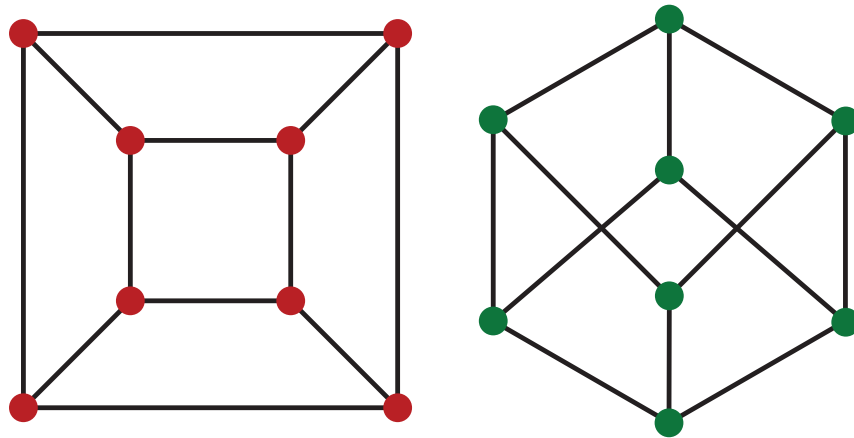


Figure 4.9 Two isomorphic graphs

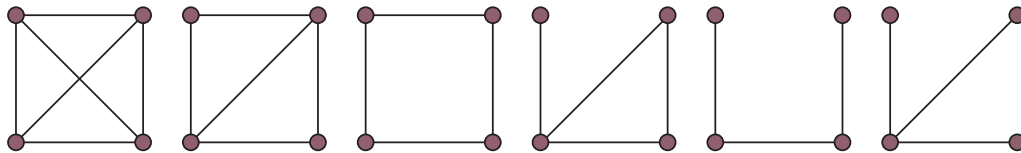


Figure 4.10 Connected simple graphs with four vertices

subgraph

Definition 4.1.27 Let Γ_1 and Γ_2 be graphs with vertex sets V_1 and V_2 . We say that Γ_1 is a subgraph of Γ_2 if there is a one-to-one map $f : V_1 \rightarrow V_2$ so that for all vertices $v_1, v_2 \in V_1$ that are adjacent in Γ_1 , also $f(v_1)$ and $f(v_2)$ are adjacent in Γ_2 .

Every graph is a subgraph of the complete graph with the same number of vertices.

tree
leaf

Definition 4.1.28 A tree is a simple, connected graph without circuits. A leaf is a vertex of degree 1 in a tree.

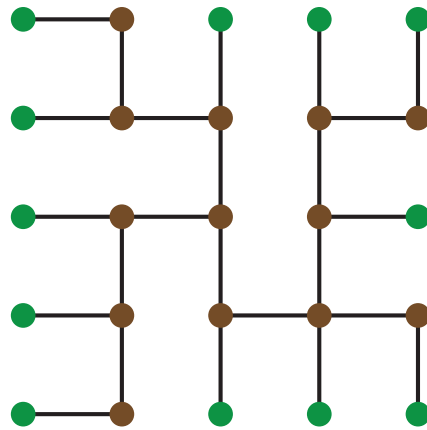
Exercise 4.1.29 Draw a tree with 25 vertices on graph paper like the one in figure 4.11 so that the vertices are at the intersections of a 5×5 grid and the edges are part of the grid, and such that no vertex has degree 2.

Solution: See figure 4.12.

Exercise 4.1.30 Determine all non-isomorphic trees with at most 7 vertices.

Solution: There are 1, 1, 2, 2, 3, 6, 11 trees with 1, 2, 3, 4, 5, 6, 7 vertices.

Exercise 4.1.31 Let Γ be a tree with v vertices and e edges. Show that $v = e + 1$.



A 5x5 grid graph with red nodes and black edges, representing a square lattice. The nodes are arranged in a regular grid, and the edges connect adjacent nodes horizontally and vertically.

Solution: If the tree has no leaves, it has just one vertex and no edges, and the claim is true. Otherwise, we can trim the leaves by removing a 1-valent vertex together with its edge. This does not change the difference $e - v$, and a trimmed tree is still a tree. We keep trimming until no leaves are left, and the claim follows.

Solution: Each move reduces the number of odd vertices by two, because it reduces the valency of the leaves to 0, and keeps the parity of all other vertices the same. The game continues until no leaves are left, but then also no edges are left. Hence, the number of moves is always the number of odd vertices divided by two.

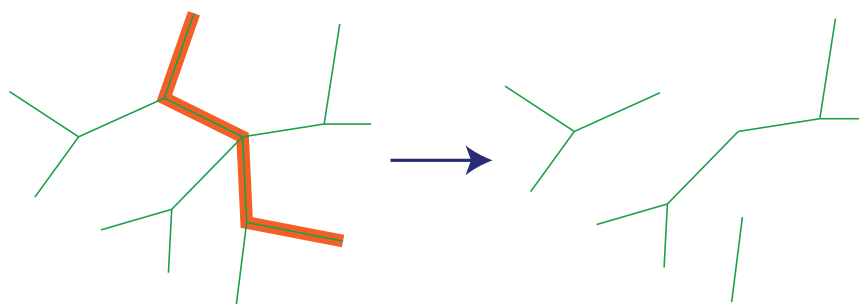


Figure 4.13 Pruning a tree

4.2 RIVER CROSSINGS

There is a popular class of puzzles where the task is to move a group of people, animals, and objects across a river, usually involving a boat that carries only a certain number in the group, and with further constraints that forbid certain participants from being left alone.

The most famous ones are probably the following ones below.

Example 4.2.1 A farmer ends up, for whatever reasons, on one bank of a river with a wolf, a goat, and a cabbage. To cross the river, the only means is a boat that can carry the farmer together with either wolf, goat, or cabbage. If left unsupervised, the wolf will devour the goat, and the goat will savor the cabbage, which the farmer won't tolerate. How does she get everything across the river?

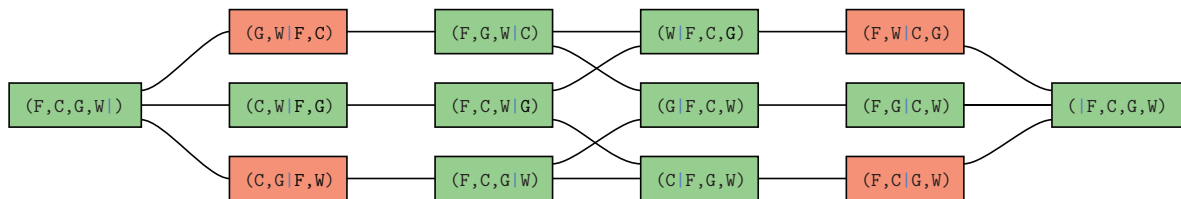


Figure 4.14 Crossing the river

It takes seven crossings. This can be found by trial and error, but a systematic treatment will show us ways to generalize the problem. In **Figure 4.14**, we have listed all possible ways to distribute the farmer with her companions on the two sides of the river. The position $(F, G | C, W)$ is an abbreviation for having the farmer with the goat on the left side of the river and the cabbage with the wolf on the right side. Such a constellation is marked red if it leads to the unfortunate demise of either cabbage or goat, and green otherwise. Two constellations are connected by an edge if a river crossing changes one into the other. Each such crossing needs to involve the farmer and one companion to be meaningful.

We see that the left and right ends of the graph represent the starting and ending positions of the problem, and a solution consists of a path from the left to the right that only visits the green constellations. This is easily possible in seven moves, and there are even two such solutions.

One way to make the problem slightly harder and to eliminate the double solution would be to stipulate that the right side of the river is infested with hungry rabbits so that the cabbage cannot be left alone by itself on that side of the river.

Exploration 4.2.2

1. What happens with two cabbages?
2. What happens with farmer, dog, cat, mouse, and cheese?
3. What happens if there are 100 objects, labeled A_1 through A_{100} , and it is known that A_i will eat A_{i+1} when left alone?
4. Make up at least one new problem of this kind by yourself, come up with a neat verbal description, and provide the solution.
5. Make up a problem that involves two boats.
6. Make up a problem that involves an island in the river, possibly of limited size.

In the previous example, there was a master whose supervision was essential, and who was the only one who could navigate the boat. In the next example, all participants can use the boat:

Exercise 4.2.1 Three missionaries and three cannibals all want to cross a river. Unfortunately, if there are more cannibals than missionaries together, temptation overwhelms the cannibals, and they will have a feast with the unlucky missionaries. There is a boat that fits just two people. Draw a graph with all constellations without fatalities as vertices, connecting two if a river crossing is possible to transform one into the other.

Solution: In this problem, both cannibals and missionaries can navigate the boat. To represent the problem in a graph, we label each vertex with the number of missionaries and cannibals on the north side of the river. For instance, the notation $3 - 2$ means that there are three missionaries and two cannibals at the north side, and consequently one missionary and two cannibals at the south side of the river. This is fatal, of course. There are only 10 non-lethal possible configurations, shown in **Figure 4.15**. We have arranged them in a coordinate grid.

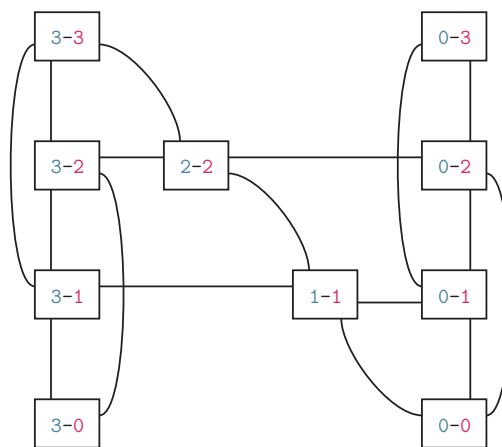


Figure 4.15 Cannibals and Missionaries

Assuming for the moment that the boat is at the north shore, a the group can move by one or two units down or right, or by one unit right and down in this coordinate grid, provided that the resulting coordinate is non-lethal again. Similarly, if the boat is at the south shore, we can move up and left by the same amounts.

We start at the top left corner and have to reach the bottom right corner, moving along the edges, but alternating between moving in the right/down and left/up direction, as the boat switches forth and back between north and south shore.

The problem is to get across from the left to the right side of the graph, and the only way to do so is to follow the path $3 - 2, 3 - 0, 3 - 1, 1 - 1, 2 - 2, 0 - 2, 0 - 3, 0 - 1$. For there beginning and the end there are two possibilities each. For instance, we can get from $3 - 3$ to $3 - 2$ in an even number of steps either via $2 - 2$ or via $3 - 1$. In either case, the complete solution takes 10 steps.

Exercise 4.2.2 Three grown-ups and two children arrive at our river. This time, the available boat is owned by a fisherman who is willing to lend the boat to the group, but only if it is returned to him at the end of the maneuver. Also, the boat can only carry one grown-up or two children. Both grown-ups and children can navigate the boat. How does the group get to the other side and return the boat to the fisherman?

Solution: In four crossings, the children can get one adult across as follows:

1. Both children cross.
2. One of them returns.

3. One adult crosses.
4. The other child returns.

This is repeated four times, so that all adults, including the fisherman, are across. Then, in the last (seventeenth) crossing, both children cross the river together.

Exercise 4.2.3 *Three humans, one big troll, and two small trolls would like to cross a river. As usual, there are a few problems:*

1. *Only humans and the big troll can navigate the boat.*
2. *The trolls can safely stay alone in a group, but if humans are present, there must be at least as many humans as trolls. Otherwise, the trolls do nasty things to the humans.*
3. *The boat only has room for two.*

Find a way for the group to safely cross the river!

Exercise 4.2.4 *You must transfer all the people (mom, dad, 4 kids, policeman, thief) across the river in this river crossing game, respecting the following rules:*

1. *The ferry can carry no more than 2 people.*
2. *Only the Adults (mom, dad, and policeman) can operate the ferry.*
3. *Dad can not be in the presence of the girls without Mom*
4. *Mom can not be in the presence of the boys without Dad*
5. *The thief cannot be alone with any of the family without the policeman.*

Solution:

1. Police Officer, Thief
2. Police Officer
3. Police Officer, Boy 1
4. Police Officer, Thief
5. Dad, Boy 2
6. Dad
7. Dad, Mom
8. Mom
9. Police Officer, Thief
10. Dad
11. Mom, Dad
12. Mom
13. Mom, Girl 1
14. Police Officer, Thief
15. Police Officer, Girl 2

16. Police Officer

17. Police Officer, Thief

The following classical puzzles make some old-fashioned assumptions about male and female behavior.

Exercise 4.2.5 *A number n of couples would like to cross a river. There is a boat seating at most p people. No husband is allowed in the presence of another woman when his wife is absent. Find a way to cross in the following cases:*

1. $n = 3, p = 2$.

2. $n = 5, p = 3$.

3. n arbitrary, $p = 4$. *In this case, it is stipulated that whenever two people of the opposite sex are together, their respective spouses must be present as well.*

4.3 EULER CIRCUITS



Figure 4.16 A stamp celebrating Euler from the German Democratic Republic

LEONHARD EULER (1707-1783) was born in Basel (Switzerland), and moved to St. Petersburg to become a professor. Following an invitation from Frederick the Great, he stayed 25 years in Berlin, but later returned to Russia. He published about a thousand articles and books in all areas of Mathematics and Physics. Half of his works were written while he was totally blind.

We now discuss the first problem in graph theory. While LEONHARD EULER was walking the streets of Königsberg in 1735, he wondered whether it was possible to follow a closed path that crosses the seven bridges of the inner city exactly once.

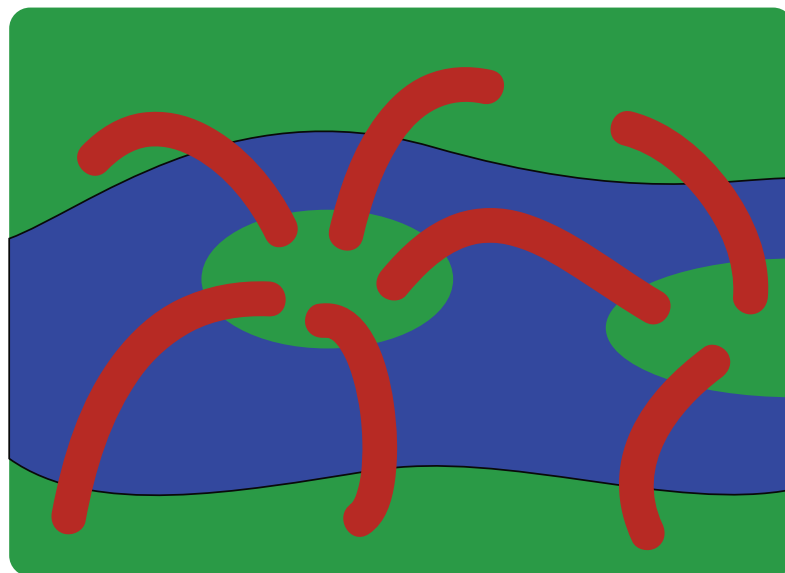


Figure 4.17 The Bridges of Königsberg

circuit

Definition 4.3.1 A circuit is a path where the first and last vertex coincide. An Euler circuit is a circuit that visits all vertices of the graph and uses all edges precisely once. Vertices may be visited multiple times.

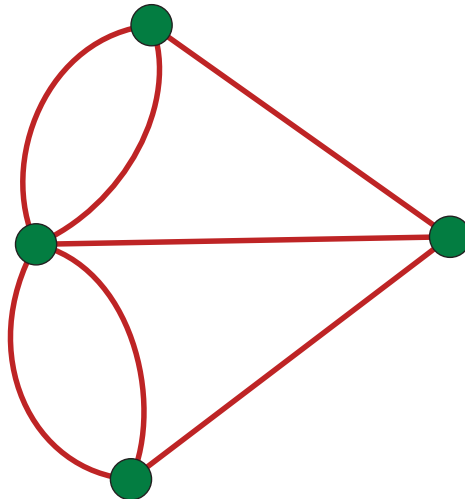


Figure 4.18 Bridges become edges

According to these definitions, Euler's bridge problem is equivalent to finding an Euler circuit in the graph **4.18**. Euler completely analyzed the problem and summarized his findings in the first theorem in graph theory. In particular, he found that the Königsberg bridge problem has no solution!

Theorem 4.3.2 *A connected graph has an Euler circuit if and only if all vertices have even degree.*

Proof. Suppose first that a graph has an Euler circuit. When we follow that circuit, we leave every vertex as often as we enter. The degree of a vertex is the number of edges we use to enter the vertex plus the number of edges we use to leave the vertex. As all edges are used, and every edge is only used once, the degree of each vertex is even.

Now, assume that every vertex of our graph has an even degree. To construct an Euler circuit, start at any vertex and move to adjacent vertices, using each edge at most once. As there are only finitely many edges available, this process will end, and our path will stop at some vertex.

We claim that our path ends at the vertex we started at. If this were not the case, we would have moved into this vertex one time more often than away from it, so that an odd number of edges at this vertex had been used. But the valency of every vertex is even, so our supposedly last vertex would still have another edge, and we could continue our path. This contradiction shows that we have produced a closed circuit that uses all edges at most once.

This circuit might not yet be an Euler circuit — we might have missed some edges. If so, there will be an unused edge ending at some vertex of our circuit; otherwise, the graph would be disconnected. Starting at this vertex, and using that edge, we follow the same procedure as before, just not using the edges of our first circuit. This results in a second circuit that shares a vertex with our first circuit. Therefore, we can combine the two circuits into a single, larger circuit.

We now keep going like this, enlarging the circuit until all edges have been used. □

Exercise 4.3.3 *For which m, n does the complete bipartite graph $K_{m,n}$ have an Euler circuit?*

Exercise 4.3.4 A triangle is a circuit with 3 vertices. We do not distinguish triangles that differ by a different starting point (like $v_1v_2v_3v_1$ and $v_2v_3v_1v_2$) or by a different sense of orientation (like $v_1v_2v_3v_1$ and $v_1v_3v_2v_1$). Show that the number of triangles in a graph with adjacency matrix A is given by $\frac{1}{6}(\text{tr}(A^3))$, where tr denotes the trace of a matrix, i.e. the sum of the diagonal elements.

4.4 MAPS AND EULER'S THEOREM

In this section, we will discuss a surprising statement about the number of vertices, edges, and faces of a polyhedron, and, more generally, of maps in the plane. It marks the beginning of a mathematical discipline called *topology*, which is concerned with features of shapes that remain unchanged if we deform the shape by squeezing and stretching.

To formulate the first version of this theorem, we need a concept from geometry:

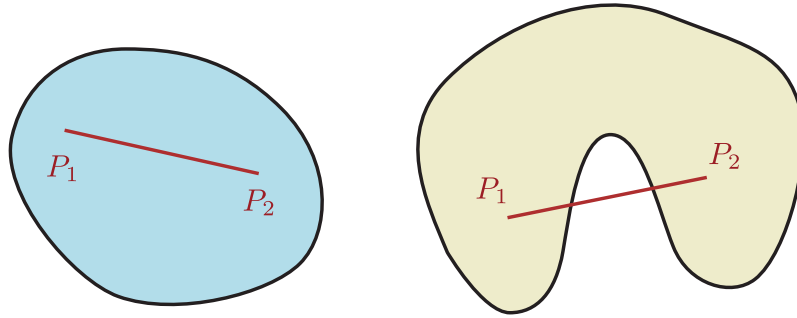


Figure 4.19 Convex vs. non-convex in the plane

convex

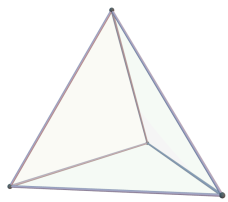
Definition 4.4.1 A subset of Euclidean space is called *convex* if for any two points in the set, the straight segment joining the points also lies within the set.

Then we have:

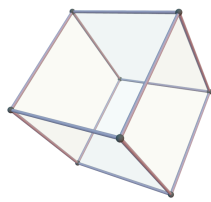
Theorem 4.4.2 Let P be a convex polyhedron with v vertices, e edges, and f faces. Then

$$\chi = v - e + f = 2$$

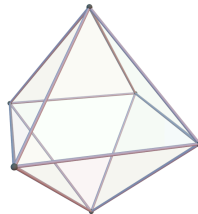
Let's look at a few examples: The *Platonic Solids* are convex polyhedra where all faces are regular polygons of the same type, and where all vertices have the same valency. There are just five of them (as we will prove in [Theorem 4.4.14](#)), shown in [Figure 4.20](#).



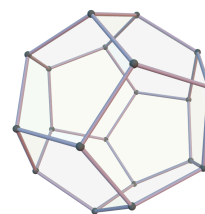
Tetrahedron



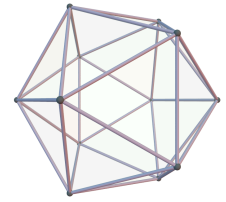
Cube



Octahedron



Dodecahedron



Icosahedron

Figure 4.20 The Platonic Solids

Exercise 4.4.3 Count the number of vertices, edges, and faces for the Platonic solids, and verify Euler's theorem for them.

It turns out that Euler's theorem statement has not much to do with convex polyhedra at all. It applies as well to maps:

map
planar
graph

Definition 4.4.4 A map is the drawing of a connected graph where the edges do not intersect. A graph that can be drawn as a map is called a planar graph.

The edge graph of any convex polyhedron is planar. To see this, we put the polyhedron into a larger sphere and project all faces, edges, and vertices radially outward onto the surface of the sphere. Then we pretend the sphere is made of rubber, and puncture it somewhere in the middle of a projected face. Finally, we pull and stretch the piece of rubber into a plane, without overlaps. This has, of course, cruelly deformed the polyhedron, but the number of edges, vertices, and faces is the same as before: The faces have become countries, the edges borders, and the vertices corners. Now we can formulate the more general version of Euler's theorem for maps:

Theorem 4.4.5 Let M be a connected planar map with v vertices (corners), e edges (borders), and f faces (countries). Then

$$v - e + f = 2$$

There are many different proofs of this theorem, and they usually require some degree of abstraction. RADEMACHER and TOEPLITZ have found an intuitive version of one of these proofs (which appears in their highly recommendable book *The Enjoyment of Mathematics*) [8], which we will use here.

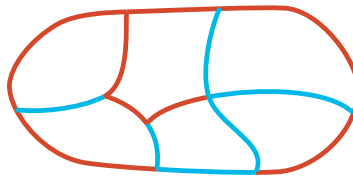


Figure 4.21 A simple map

Proof. For illustration purposes, here is an example of a map with 7 countries (don't forget to count the outside region, too, as a country), 14 borders, and 9 corners. Ignore the colors for the moment.

Imagine the large outside country is in fact the sea, and the inner countries are fields that are protected from the water by dams (which are the borders). The fields are usually used for cattle and need to be flooded occasionally so that they are exposed to fertilizing saltwater.

We first ask: How many of the dams (borders) need to be breached at least so that all countries are flooded? By cutting down one dam, we can flood at most one field. On the other hand, if a field is still dry, we can always find a dam to cut down so that a (possibly different) dry field is flooded. Therefore, the number of dams that need to be removed is exactly $f - 1$, the number of initially dry fields.

The next step is to analyze the network of remaining dams. We claim that one can walk from any point on this network to any other point without getting wet. If this were not the case, there would be two islands of dams separated by water, which were formerly connected by dams. But this means we could have left some of the dams that connected the islands still intact, while still flooding all the fields.

Secondly, we claim that all corners of the original dam network are still part of the remaining dams. If not, we would have removed all the dams ending in one corner. But then we could have left any one of these dams intact and still flooded all the fields.

Thirdly, we claim that from any point to any other point, there is only one path along the dry dams. Otherwise, we would have a closed circuit of dams, which would protect at least one field.

These three claims allow us to count the number of dams that are in the network. We continue removing dams by choosing at each step a dam that ends in a corner. Each such step will remove one dam and one corner from our network, and the remaining network will still have no closed circuits and be connected. This process continues until we are left with only one dam, which has two corners. Therefore, the number of dams in our remaining network is exactly $v - 1$.

The two counting arguments together show that the total number of dams is

$$e = (f - 1) + (v - 1)$$

which is Euler's theorem.

Corollary 4.4.6 *For a tree with v vertices and e edges, we have $v = e + 1$.*

Exercise 4.4.7 *Suppose a convex polyhedron has only hexagons and triangles as faces. Assume further that all vertices are 3-valent. How many triangles does the polyhedron have? Find an example of such a polyhedron.*

Solution: Let f_3 and f_6 be the numbers of triangles and hexagons in the polyhedron, and $v_a = 4$ and v_b the number of vertices surrounded by three hexagons and by two hexagons and one triangle, respectively. Also let v be the number of vertices, $f = f_3 + f_6$ the number of faces, and e be the number of edges. As the vertices are all 3-valent, we get $2e = 3v$. We also get from the face count $2e = 3f_3 + 6f_6$, so that by Euler

$$\begin{aligned} 2 &= v - e + f \\ &= f - \frac{1}{3}e \\ &= f - \frac{1}{3}e \\ &= f_3 + f_6 - \frac{1}{6}(3f_3 + 6f_6) \\ &= \frac{1}{2}f_3 \end{aligned}$$

so that

$$f_3 = 4.$$

Exercise 4.4.8 *Show that any 4-valent convex polyhedron containing only triangles and squares has precisely 8 triangles as faces. Find two examples of such polyhedra.*

Solution: Let v , e , and f be the number of vertices, edges, and faces, respectively. Then $2e = 4v$ by the valency condition. From Euler, we know that $2 = v - e + f$, so the previous equation tells us that $2 = f - \frac{1}{2}e$. Let f_3 and f_4 be the number of triangles and squares. Then $f_3 + f_4 = f$ and also $2e = 3f_3 + 4f_4$. We use these equations to replace f and e in $2 = f - \frac{1}{2}e$. This gives

$$2 = f_3 + f_4 - \frac{1}{2}\left(\frac{3}{2}f_3 + 2f_4\right) = \frac{1}{4}f_3$$

This implies $f_3 = 8$.

Exercise 4.4.9 A convex polyhedron has only hexagons and squares as faces. All vertices have degree 3. How many squares does the polyhedron have?

Exercise 4.4.10 Use Euler's formula to show that a trivalent convex polyhedron containing only hexagons and pentagons has precisely 12 pentagonal faces. Examples of such polyhedra are the dodecahedron and the soccer ball.

Exercise 4.4.11 Show that a 4-valent convex polyhedron containing only triangles and squares has precisely 8 triangles as faces. Give at least two examples of such polyhedra.

Exercise 4.4.12 A map has v vertices, e edges, and f faces, including the outside face. All faces are either 4- or 5-gons, and all vertices have either degree 3 or 4.

1. Show that any such map has eight more vertices of degree 3 than 5-gons as faces.
2. Find an example of such a map with just one 5-gon, or show that it cannot exist.
3. Find an example of such a map with two 5-gons and as few 4-gons as you can.

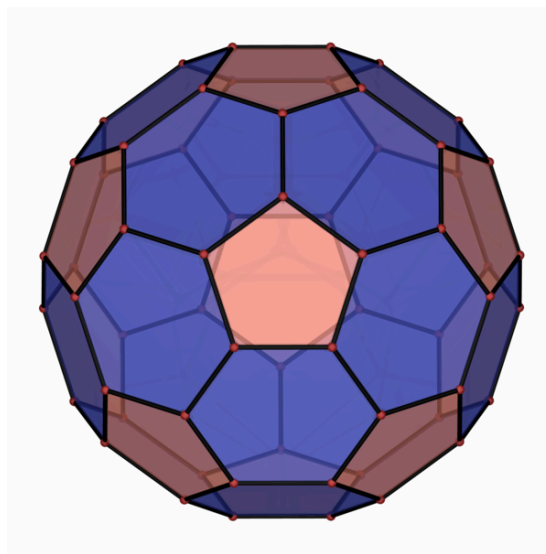


Figure 4.22 The truncated icosahedron, or soccer ball

As an important application of Euler's formula, we show that the only Platonic solids are those from our list. We will do this by analyzing regular planar maps.

regular
map

Definition 4.4.13 A planar map is called **regular** if all vertices have the same degree and all faces have the same number of vertices.

Theorem 4.4.14 *Let M be a regular planar map. Then M is the map of a tetrahedron, cube, octahedron, dodecahedron or icosahedron.*

Proof. To see this, let v, e, f be the number of vertices, edges, and faces of M as usual. Let n be the number of edges per vertex and m be the number of edges per face. Then

$$2e = nv = mf$$

and by Euler's formula,

$$2 = v - e + f = \left(\frac{2}{n} - 1 + \frac{2}{m}\right)e$$

or

$$\frac{1}{e} = \frac{1}{m} + \frac{1}{n} - \frac{1}{2}$$

Now m and n are at least 3 and can't be both greater than 4. Inspection shows that the only possible cases for m and n are

$$(m, n) = (3, 3), (3, 4), (4, 3), (3, 5), (5, 3)$$

Observe that these five possibilities determine the vertex, edge, and face counts for the Platonic solids, by the computation above. What remains to be done is to show that in each of the above cases, the Platonic map is the only map with the same count.

This is true, but requires either an argument from topology or some tedious discussions (like trying out all possibilities), and we will omit this. The argument from topology uses the fact that the sphere is *simply connected* and the theory of covering spaces.

Exploration 4.4.1 The Platonic maps give examples of maps where an n -gon is subdivided into n -gons.

1. Can you find other examples of such subdivisions?
2. For which numbers m and n can you divide an m -gon into n -gons?
3. Suppose you have an m -gon within another m -gon. Can you divide the region between the two m -gons into n -gons?
4. Can you show that for some m and n such subdivisions are impossible?

As another application of Euler's theorem, we will show that certain graphs are *not* planar.

An example is given by the following puzzle: Three old and very stubborn triangles and three old squares live in different places on a large meadow. The triangles hate each other as badly as the squares, but all triangles love all squares, and vice versa. So each triangle would like to have a path to each square, and no two paths should cross. Can this be achieved?

In other words, we would like to know whether the bipartite graph $K_{3,3}$ is planar. The answer is no, and Euler's theorem will help us to find a convincing argument.

We begin with a simple observation:

Lemma 4.4.15 *A map with v vertices, all of whose faces are triangles, has $3v - 6$ edges.*

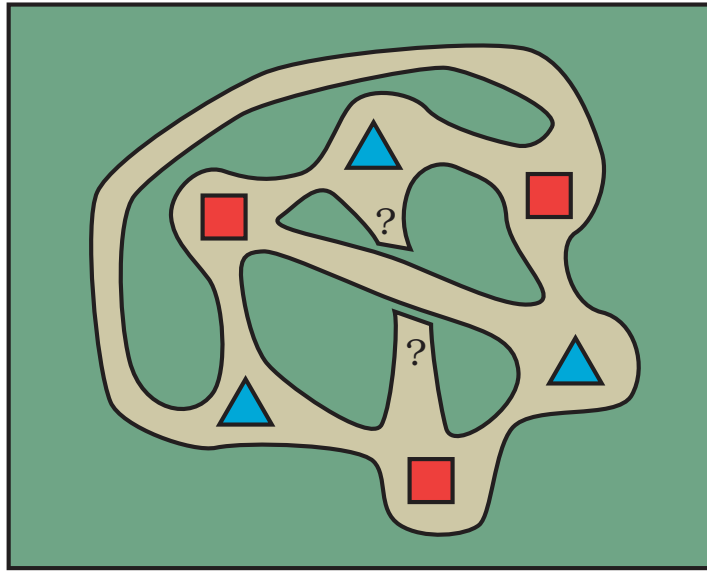


Figure 4.23 The $K_{3,3}$ dilemma

Proof. Because each face is triangular, we have $3f = 2e$. Thus, by Euler,

$$\begin{aligned} 2 &= v - e + f \\ &= v - e + \frac{2}{3}e \\ &= v - \frac{1}{3}e \end{aligned}$$

Solving for e proves the claim. □

Corollary 4.4.16 *A map with v vertices can have at most $3v - 6$ edges.*

Proof. Any map can be turned into a triangular map by just adding edges. Because we just proved the claim for the augmented map, it must be true even more so for the original map which has fewer edges. □

Corollary 4.4.17 *Neither K_5 nor $K_{3,3}$ is planar, nor can they be contained in a planar graph.*

Proof. Suppose we could realize K_5 as a planar map. As it has 5 vertices, it could have at most 9 edges. It has, however, 10.

The argument is not quite as simple for $K_{3,3}$, as it has 6 vertices and 9 edges. However, as it is bipartite, every cycle is of even length, and in particular, each face has an even number of edges. Similar to the above reasoning, we have $4f \leq 2e$, so that because of Euler, $2 = v - e + f \leq v - \frac{1}{2}e$, or $e \leq 2v - 4$, which leads to a contradiction for $K_{3,3}$. □

There is an important and difficult theorem by KAZIMIERZ KURATOWSKI that characterizes planar graphs in an elegant way:

Theorem 4.4.18 *A graph without 2-valent vertices is planar if and only if it does not contain a subgraph that arises from K_5 or $K_{3,3}$ by subdividing edges.*

4.5 COLORING GRAPHS AND MAPS

A common problem in graph theory asks to assign a color to each vertex so that adjacent vertices have different colors.

Definition 4.5.1 The chromatic number of a graph is the smallest number of colors that are needed to color its vertices so that adjacent vertices have different colors.

chromatic
number

Example 4.5.1

1. A graph is bipartite if and only if its chromatic number is 2.
2. The chromatic number of the complete graph K_n is n .

Exercise 4.5.2 Find a graph with as few vertices as possible with chromatic number 5 and where all vertices have valency at most 3.

Exercise 4.5.3 Let $n > 1$ be an integer and V_n the set of all n -tuples containing only 0 and 1. Thus V_n has 2^n elements. Connect two n -tuples by an edge if and only if they differ in exactly one entry. For instance, $(0, 1, 0, 1)$ and $(1, 1, 0, 1)$ are connected, but not $(0, 1, 0, 1)$ and $(1, 1, 1, 1)$. The resulting graph is called the n -cube. Determine the chromatic number of the n -cubes for $n = 2, 3, 4, 5$.

We next turn to coloring maps. This refers to coloring the faces of the map, not the vertices of the corresponding graph. So it looks as if map coloring and graph coloring are two different things. This is, however, not the case.

Besides the edge graph of a map, there is a second way to associate a graph with a map, which uses the concept of the *dual map*.

Definition 4.5.4 For a map, we construct its dual map as follows: In each face, we pick a point which will become a vertex of the dual map. Two of these new vertices are connected by an edge if the two faces are adjacent, i.e., share an edge. Finally, the faces of the dual map correspond to the vertices of the old map.

dual
map

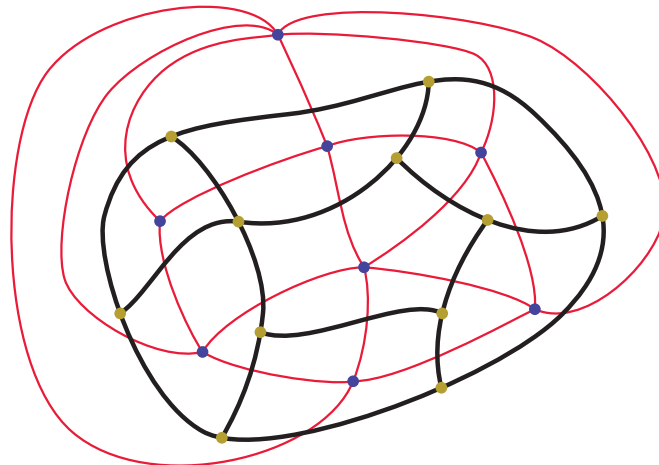


Figure 4.24 The dual map of a map

This way, the face coloring of a map becomes the vertex coloring of the edge graph of the dual map.

One of the most famous theorems in graph theory is the following:

Theorem 4.5.5 *Any map can be 4-colored.*

The original proof by ALFRED KEMPE in 1879 contained a serious gap which was discovered only 1890 by PERCY JOHN HEAWOOD. It then took until 1977, when KENNETH APPEL and WOLFGANG HAKEN gave a computer-supported proof. Subsequent work has simplified and streamlined this proof, but there is still no proof available that a human would want to read.

We are content with proving the weaker

Theorem 4.5.6 *Any map can be 5-colored.*

We will divide the argument into smaller pieces for better digestion.

First, we will show that we can make several simplifying assumptions.

For instance, we can assume that our map is connected: Otherwise, we can connect the components by simple edges (called *bridges*). If the new graph can be colored with n colors, the same holds for the old graph.

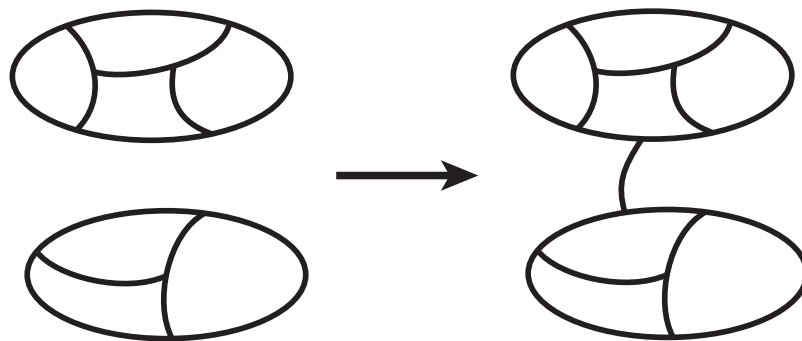


Figure 4.25 Making a graph connected

We can also assume that the map doesn't contain loops: Suppose our map has a loop. This loop separates the map into an inside and an outside map with the loop as the only common border. By induction over the number of faces, we can assume that we are able to color the inside and outside separately. Then we only need to change the colors of (say) the inside map so that the colors of the two faces bordering the loop become distinct.

Next, we can assume that the map doesn't contain *bridges*, i.e., edges that make the graph disconnected when we remove them. That's because any graph with a bridge can be simplified by contracting the bridge to a single vertex, without the necessity to change the coloring.

Finally, we can assume that there are no vertices with degree 2 or 1: Vertices of degree 1 can be removed together with the edge they belong to without changing the coloring, and vertices of degree 2 can be removed as well.

Let's call a connected graph without vertices of degree 1 or 2, and without bridges or loops, a *nice* graph. We have shown that

Lemma 4.5.7 *If all nice maps can be colored with n colors, then all maps can be colored with n colors.*

In the next step we will show

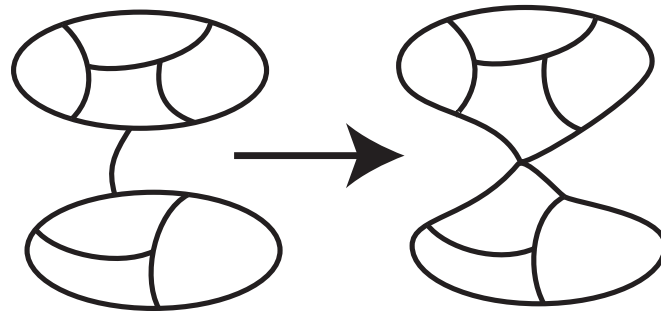


Figure 4.26 Contracting an edge

Lemma 4.5.8 *A nice map has at least one face that shares a border with at most five other faces.*

Proof. To see this, we proceed by contradiction: Assume that we have a nice map where all faces border at more than five faces. Denote by v, e, f the number of vertices, edges, and faces of the map.

Let f_i be the number of faces that have precisely i edges as borders. As we have no loops, $f_1 = 0$. Clearly

$$f = f_2 + f_3 + f_4 \cdots$$

but also

$$2e = 2f_2 + 3f_3 + 4f_4 + \cdots$$

because if we count the number of edges by summing over all edges that belong to each face, we count every edge twice.

Now we evaluate

$$\begin{aligned} \sum_{i \geq 2} (6 - i) f_i &= 6 \sum_{i \geq 2} f_i - \sum_{i \geq 2} i f_i \\ &= 6f - 2e \\ &= 6(2 + e - v) - 2e \quad \text{by Euler's theorem 4.4.2} \\ &= 12 + 4e - 6v \\ &\geq 12 \quad \text{by the Handshaking Lemma 4.1.15} \end{aligned}$$

Because the summands $(6 - i) f_i$ are ≤ 0 for $i > 6$ and the sum is positive, at least one of the numbers f_2, f_3, f_4 or f_5 must be positive. $\square$

Now we can prove the 5-color theorem:

Proof. Let's assume that the 5-color theorem is false. Then there exists a nice map that cannot be colored with five colors and that has as few faces as possible for such a map. By **Lemma 4.5.8**, this map has a special face with only five or fewer neighbors.

We first show that we can color all of the map except for this face with five colors. To see this, we remove this face from the map by dragging all of its vertices to the center, thereby enlarging the surrounding faces. The result is a nice map with fewer faces, which can be 5-colored, because our original map was the one with the smallest number of faces that couldn't be colored with five colors. Using the same coloring for the old map colors it with five colors, except for the special face with five or fewer neighbors, which we leave white.

If the neighbors of the special face use only four or fewer colors, we can color the special face with an unused color by the neighbors, and we are done. Thus, the only case to consider is that we have an uncolored (white) face with five bordering faces, all colored in different colors, as indicated in **Figure 4.27**.

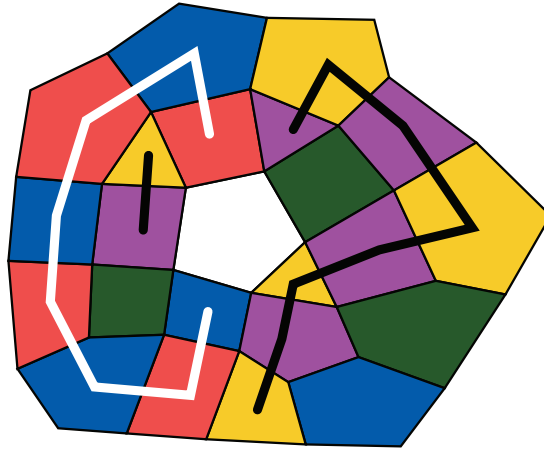


Figure 4.27 Alternating chains

These faces come in a natural counterclockwise order. Let's look at two of these neighbors that are not next to each other along the border of the white face, like the red and blue faces.

It happens so in the example that these two faces can be connected by a chain of adjacent faces that alternate in color. This red-blue chain is marked by a white path.

Let's now consider a second pair of differently colored faces that border on the white face but are not consecutive and separated by the first chain. These could be yellow and purple. There cannot be a yellow-purple chain connecting these two faces, as this chain would cross the red-blue chain. This is not possible, as a face can only have one color. Thus, no matter what, we find a pair of non-consecutive neighbors that a color-alternating chain cannot link. Take one of these two faces, and connect it to all possible faces by color-alternating chains that use only the two colors. Then swap the colors in the chain. The resulting coloring is still a valid 5-coloring of the map, but now we only use 4 colors for the neighboring faces of the white face. This allows us to color the white map with the color we haven't used, and we are done. $\square$

Exercise 4.5.9 *Somebody has drawn a finite set of circles in the plane. Show that the resulting map can be colored with just two colors.*

4.6 BRUSSEL SPROUTS

5 Solitaire Games

5.1 PEG SOLITAIRE

Peg solitaire is usually played on a cross-shaped board with circular holes as shown, where all holes except the center one are filled with pegs. A move consists of jumping with one peg across another one horizontally or vertically into an empty hole and removing the peg that has been jumped over. The goal is to remove all but one peg, and to have the last remaining peg end up in the center hole.

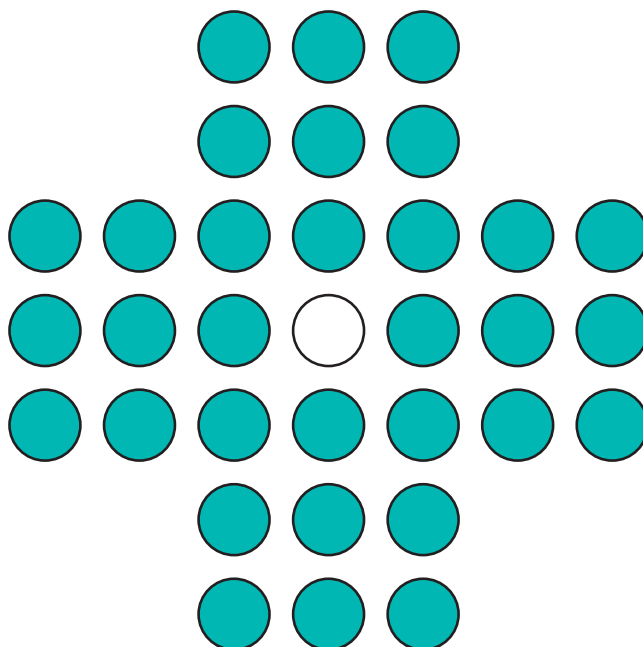


Figure 5.1 Peg Solitaire on a standard board

As there are 33 pegs, a solution will take 32 jumps. However, if one counts consecutive jumps of the same peg as a single move, much shorter solutions are possible. A solution in 18 moves was found in 1912 by ERNEST BERGHOLD (see Figure 5.2), and 1964 JOHN BEASLY proved that there is no shorter solution.

There are many variations to this puzzle: One can consider other starting and ending positions or different board layouts for instance. We will now discuss how group theory helps to determine how the location of the initial empty space determines the location of the last peg. To this end, we color the board with three colors as in Figure 5.3.

Observe that any single move removes pegs of two distinct colors and adds a peg of the third color. If instead of colors we use the binary numbers 01_2 , 10_2 and 11_2 , a move will not change the NIM-sum of the values of the positions that are occupied by pegs. The NIM-sum is reviewed in Table 5.1

In other words, the NIM-sum of all values of the occupied holes in a position does not change during the game. We call this sum the *value* of the position

To compute the value of the initial position, note that a fully occupied board has value 0, as all colors show up equally often. Hence, a board position has the same value as the *inverted* board, where all unoccupied holes are occupied, and vice versa.

In particular, the initial position with just the center hole empty has the same value as position with just the center hole occupied, which is 10. Consequently, all positions that can be obtained from the initial position must have value 10, and the only possible end positions with a single peg must have that peg in a hole marked 10. We can narrow the possibilities down a bit further by exploiting that the board is more symmetric than its coloring by NIM values: If we rotate the coloring by 90° , the center hole is still labeled 10, and thus the final peg must again end up in a hole that is

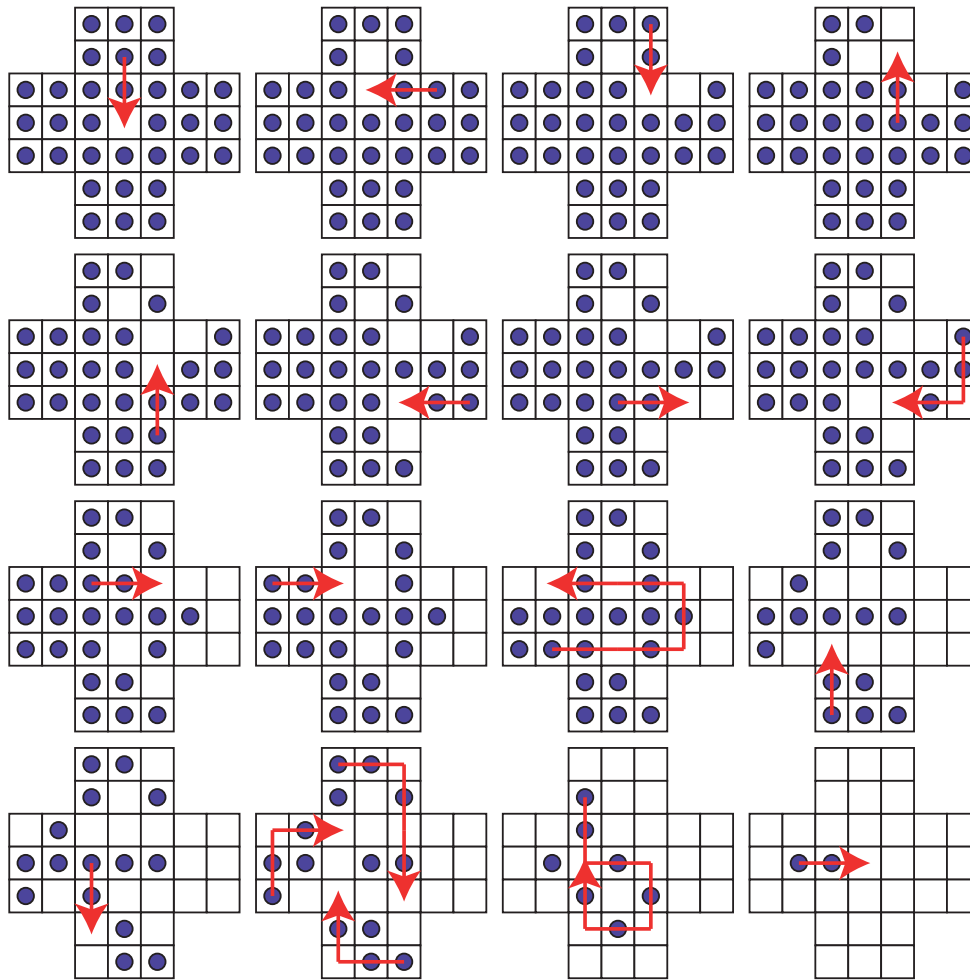


Figure 5.2 Berghold's solution in 18 moves

now marked 10 as well. This leaves essentially two different possibilities for the peg to end: In the center square, as asked for in the original puzzle, or in one of the four middle holes of the far edges of the square.

This argument does not guarantee that there is an actual solution for solving a particular puzzle. It only tells us that certain problems cannot have a solution at all.

Exercise 5.1.1 Find a solution for the solitaire problem with the peg ending in one of the far edges of the board.

Triangular peg solitaire is played on a triangular board where moves are now allowed in all six straight directions. Below is a board with a triangle size 5.

Exercise 5.1.2 Suppose initially all holes are occupied except for the top one. Find all possible ending positions that have only one peg on the board, and give a solution for each.

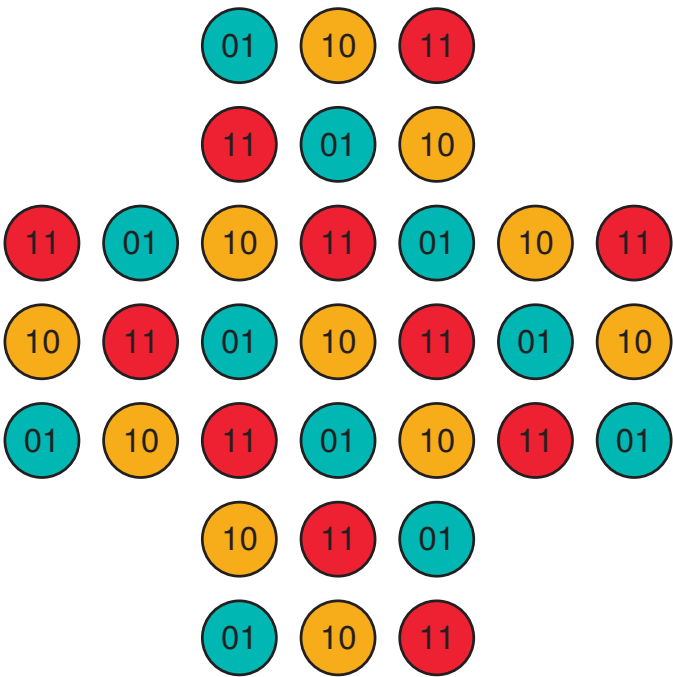


Figure 5.3 Peg solitaite board coloring

+	0	01	10	11
0	0	01	10	11
01	01	0	11	10
10	10	11	0	01
11	11	10	01	0

Table 5.1 Nim-Addition reviewed

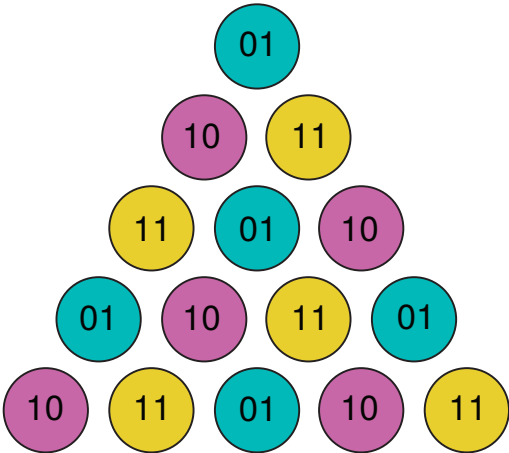


Figure 5.4 Peg Solitaire on a triangle board

5.2 PERMUTATIONS

Definition 5.2.1 A permutation of a set of objects, say the numbers $\{1, 2, \dots, 6\}$ is an arrangement of all elements in the set in a particular order, like $[2, 5, 1, 4, 6, 3]$.

How many different permutations does a set of n elements have? We claim that there are $n! = 1 \cdot 2 \cdots n$ permutations. To see this, we are going to construct all permutations as follows: For the first entry, we can choose among n elements. After doing so, there are $n - 1$ elements left for the second entry, and so on.

The set of all permutations of the n elements $\{1, 2, \dots, n\}$ is denoted by S_n . For instance, S_3 is the set $\{[1, 2, 3], [1, 3, 2], [2, 1, 3], [2, 3, 1], [3, 1, 2], [3, 2, 1]\}$.

Many puzzles are concerned with bringing a certain set of objects in some sort of order, using certain operations. Here is a simple

Example 5.2.1 Given a permutation of the first n positive integers, can you bring this in consecutive order if you are only allowed to exchange consecutive elements?

This is possible and quite easy: A step in the BUBBLE SORT algorithm parses the permuted list of numbers until a consecutive pair is found that is not in the correct order. This pair is then swapped. When no such pair exists, the list is sorted. Otherwise, the step is repeated. This algorithm works because it decreases in each step the *disorder* of the permutation. To make this precise, we make the following

Definition 5.2.2 The disorder of a permutation $\sigma = [a_i]$ of the numbers 1 to n is defined as

$$\text{dis}(\sigma) = \#\{(i, j) : i < j \text{ and } \sigma_i > \sigma_{i+1}\}$$

Clearly, a permutation is sorted if and only if its disorder is 0, and the maximal disorder is reached for the reverse permutation $[n, n - 1, \dots, 2, 1]$, where all $\frac{1}{2}(n + 1)n$ pairs with indices $i < j$ are in the wrong order. In fact, transposing two neighboring elements decreases the disorder precisely by 1, because interchanging a_i and a_j only affects the state of order of pairs with indices between i and j .

We can illustrate this by turning the set of permutations of order n into a graph, where the vertices are the elements of S_n , and two permutations are connected by an edge if and only if there is a transposition that takes one permutation to the other.

If two elements that are further apart are being transposed, the effect on the disorder is less simple, but we have the fundamental observation:

Theorem 5.2.3 A transposition changes the disorder of a permutation by an odd amount.

Proof. We can realize a transposition of the a_i with a_{i+d} by first transposing a_i successively with its d neighbors to the right, and then a_{j-1} with its $d - 1$ neighbors to the left. These are $2d - 1$ many swaps, each changing the disorder by ± 1 . $\square$

This motivates the

Definition 5.2.4 A permutation is called even or odd if its disorder is even or odd, respectively.

We then have

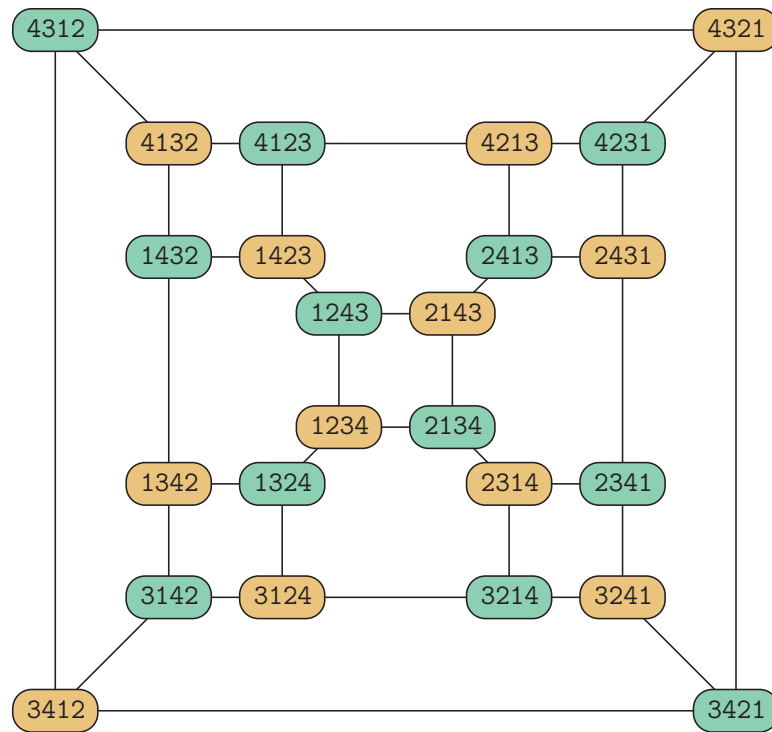


Figure 5.5 The Permutation Graph

Corollary 5.2.5 *The product of an even (resp. odd) number of transpositions is an even (resp. odd) permutation.*

Exercise 5.2.6 *For a given permutation, you are allowed to cyclically permute three consecutive elements. Can you bring any permutation in order just by applying these cyclical moves?*

Solution: A cyclical permutation of three consecutive elements can be realized with two transpositions. Therefore, each such move preserves the parity of a permutation. As a consequence, no odd permutation can be brought into order.

Exercise 5.2.7 *Here is another definition of even and odd permutations. Consider the product $\prod_{1 \leq i < j \leq n} a_j - a_i$. That is some big non-zero number. For instance, for $n = 4$ and $a = [1, 2, 3, 4]$, we get $(2-1)(3-1)(4-1)(3-2)(4-2)(4-3)$. Show that a permutation is even if and only if this product is positive.*

There is another way to interpret presentations: Instead of considering them as a collection of objects in disarray, we can also view them as instructions on how to move things. For instance, the permutation $a = [a_i]$ can be interpreted as the map $i \mapsto a_i$. As these maps are not one-to-one, a permutation of a set of elements is nothing but a bijection of that set onto itself.

In particular, a permutation has an inverse permutation that undoes the effect of the given permutation. For instance, the permutation $[3, 4, 5, 2, 1]$ has as its inverse the permutation $[5, 4, 1, 2, 3]$.

The *cycle representation* of a permutation helps to understand the effect a permutation has when considered as a map. For example, let's see to what positions the number 1 is moved to when we repeatedly apply the permutation $[2, 3, 5, 4, 1]$ to itself. First, it's moved from position 1 to 2, then to 3, to 5, and then back to 1. After that, it starts all over.

The sequence $\langle 1235 \rangle$ is called a *cycle*. It tells us where all numbers in it are being moved to: the subsequent number in the cycle. The only number we don't know about is 4, which has its own cycle $\langle 4 \rangle$. Thus the cycle representation of the permutation $[2, 3, 5, 4, 1]$ becomes $\langle 1235 \rangle \langle 4 \rangle$. If we agree that short cycles of length 1 like $\langle 4 \rangle$ don't need to be written down, we can also just write $\langle 1235 \rangle$.

Here is another example. The permutation $[3, 4, 5, 2, 1]$ has two cycles, namely $\langle 135 \rangle$ and $\langle 24 \rangle$, so its cycle representation is $\langle 135 \rangle \langle 24 \rangle$.

Cycles of length two like $\langle 24 \rangle$ are just transpositions.

If a permutation is given in cycle form, its inverse can be found just by reversing the order of each cycle. For instance, the inverse of $\langle 135 \rangle \langle 24 \rangle$ is $\langle 531 \rangle \langle 42 \rangle$.

Exercise 5.2.8 Write the following permutations in cycle form, determine their order, and find their inverse in standard row notation.

1. $[54321]$
2. $[1357246]$
3. $[214365]$
4. $[234561]$
5. $[3425176]$

Exercise 5.2.9 Write the following permutations in row notation, determine their order, and find their inverse in row notation.

1. $\langle 135 \rangle \langle 24 \rangle \langle 6 \rangle$
2. $\langle 1234567 \rangle$
3. $\langle 13 \rangle \langle 24 \rangle \langle 56 \rangle$
4. $\langle 13246 \rangle \langle 57 \rangle$
5. $\langle 162 \rangle \langle 534 \rangle$

If we have a set of permutations given, the set of all permutations that can be reached by successively applying these permutations and their inverses is called the *group* of permutations generated by the set. For instance, we have seen that the swaps (the transposition of neighbors) generates all permutations, while pairs of transpositions like $\langle 13 \rangle 24$ generate the even permutations.

Exercise 5.2.10 *For the following set of permutations, determine all elements in the permutation group generated by the set.*

1. $\{[1423], [4123]\}$
2. $\{[1432], [4231]\}$
3. $\{[54321], [12354]\}$
4. $\{[132], [231]\}$
5. $\{[2143], [4321], [3124]\}$

5.3 THE 14-15 PUZZLE

This game consists of a 4×4 square board with 15 movable plates, labelled 1 through 15 over each square. One square is left empty. In the beginning, the plates are distributed randomly, and the goal is to restore the ordered goal position shown below by applying legal moves. A legal move consists of moving one plate horizontally or vertically one unit onto the currently vacant square.

This game became enormously popular when SAM LLOYD offered a \$1,000 reward in the 1870s to solve the position that has just blocks 14 and 15 interchanged.

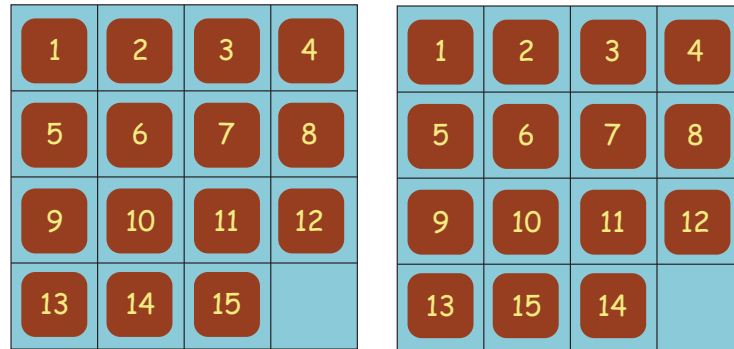


Figure 5.6 Goal position and Lloyd's \$1,000 reward position

It turns out that Lloyd's money was safe, as his problem has no solution. In fact, only half of all positions are solvable.

To analyze the game, we will proceed in two steps. We will show first that any position can be turned into either the goal position or Lloyd's position. Then, we will show that these two positions cannot be reached from one another by a sequence of legal moves.

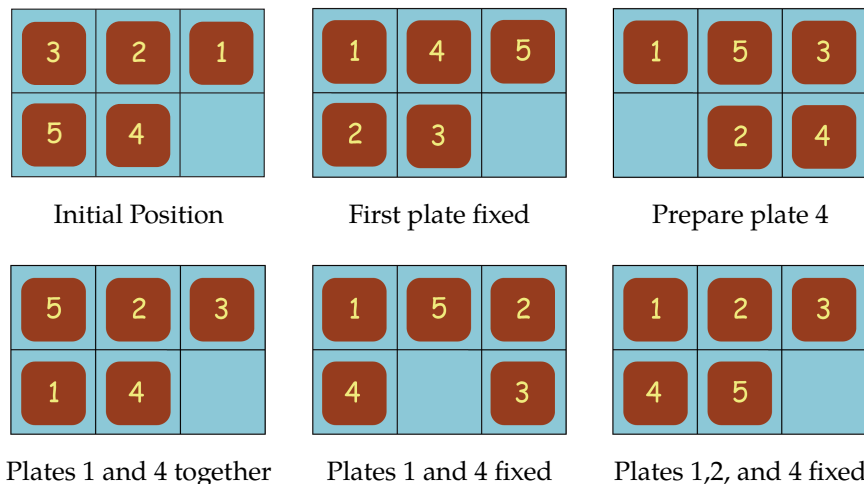


Figure 5.7 Doubling Marbles

Let us begin by studying the same puzzle on the smaller 2×3 board.

In the first step, we move plate 1 to the upper left position. This can be easily accomplished by just moving all plates cyclically. The second step will move plate 4 to the lower left position. We can assume that plate 4 is somewhere else, i.e., in the right 2×2 square. Cyclically rotate the plates in that right square until plate 4 ends up in the lower right corner. Then, move the plates except 1 and 4 so that the empty plate is below plate 1. The next goal is to bring plates 1 and 4 together. To this end, move all plates in the left 2×2 square counterclockwise once. After that, move plate 4 to the left.

In the next steps, move the empty square to the middle top position, without moving plates 1 or 4. Then move all plates in the left 2×2 square clockwise once.

Now plates 1 and 4 are in the correct position, and we can still move the plates into the right 2×2 square so that plate 2 is in the top middle position and the empty square is in the bottom right position. We can't be sure, however, that plates 3 and 5 are in their correct positions.

This essentially shows:

Proposition 5.3.1 *On a 2×3 board, the plates can be arranged so that the empty square is at the bottom right position and three of the five plates are at prescribed positions.*

Exercise 5.3.2 *Find solutions for the remaining cases of the proposition. You can, of course, use the case we did to simplify your work.*

Now we indicate how to bring a general 4×4 board into either the goal position or Lloyd's position. It is fairly easy to solve the two top rows, so that we essentially only have to deal with a 2×4 board. Using the similar moves as for the 2×3 board, we can bring the correct plates (9 and 13) into their positions on the leftmost squares. This leaves us with a 2×3 board, which we can solve up to possibly two exchanged tiles.

Next, we will learn that Lloyd's position cannot be solved.

For this, we will introduce an auxiliary plate labeled 16, which we place on the empty square. Any position now becomes a permutation of the numbers $1, \dots, 16$, and a move consists of swapping plate 16 with a horizontal or vertical neighbor. These swaps are transpositions and thus odd permutations. Now observe that for the plate 16 to return to its initial position, it takes an even number of swaps, both horizontally and vertically. This means that any sequence of moves that returns plate 16 to its starting position effects an even permutation of the plates. However, turning Lloyd's position into the goal position challenges us to accomplish an odd permutation, which is impossible!

5.4 CONWAY'S ARMY

DOUBLING MARBLES is being played in the first quadrant. Initially, marbles are placed on the squares $(0,0)$, $(1,0)$, and $(0,1)$. A move replaces a marble at position (i, j) by marbles at the positions $(i+1, j)$ and $(i, j+1)$. The goal is to empty the lower left triangle consisting of the squares $(0,0)$, $(1,0)$, and $(0,1)$.

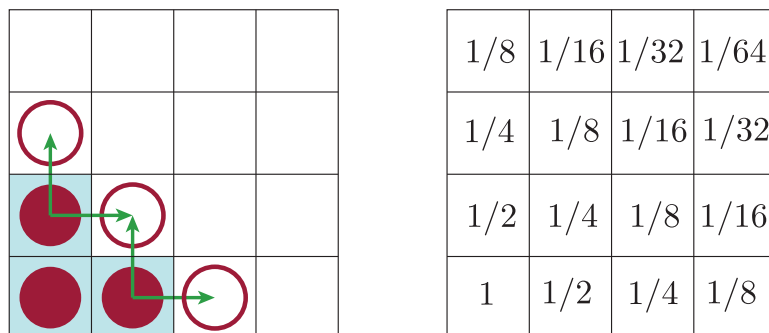


Figure 5.8 Doubling Marbles

Put a weight on each square at position (i, j) of $2^{-(i+j)}$. So the square at $(0,0)$ has weight 1, and the squares at $(2,0)$, $(1,1)$ and $(0,2)$ all have weight $1/4$.

Assign a weight to a position by adding the weights of all occupied squares. We then have the first fundamental observation:

Proposition 5.4.1 *A move does not change the weight of a position.*

Proof. A move removes a marble at position (i, j) and replaces it by two new marbles at $(i+1, j)$ and $(i, j+1)$. So the weight of the position is decreased by $2^{-(i+j)}$ but also increased by $2^{-(i+1+j)}$ and $2^{-(i+j+1)}$, which proves the claim. $\square$

The second fundamental observation is that the weight of a position can not be arbitrarily large:

Proposition 5.4.2 *If all positions in the first quadrant $\{(i, j) : i, j \geq 0\}$ are occupied, then the total weight of the position is equal to 4. In particular, any finite position in the first quadrant has total weight strictly less than 4.*

Proof. We compute

$$\begin{aligned} \sum_{i,j=0}^{\infty} 2^{-(i+j)} &= \sum_{i=0}^{\infty} 2^{-i} \cdot \sum_{j=0}^{\infty} 2^{-j} \\ &= 2 \cdot 2 = 4. \end{aligned}$$

Theorem 5.4.3 *Given the position with three marbles in the triangle $(0,0)$, $(1,0)$, and $(0,1)$, there is no sequence of moves that moves all three marbles out of this triangle.*

Proof. The total weight of the starting position is 2, while the total weight of any finite position outside the triangle is strictly less than $4 - 2$. $\square$

DOUBLING MARBLES can also be played on a triangle board, with the disks placed on the vertices of the triangles. A move is applied to a triangle that has only one vertex occupied, resulting in that triangle with the other two vertices occupied. Each move increases the number of disks by 1. Now we place a weight of $1/2^i$ on each vertex in row i . Again, the weight of a position is the sum of the weights of all occupied vertices.

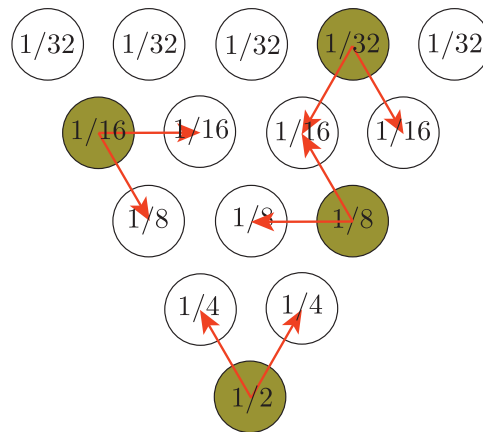


Figure 5.9 Doubling Marbles on a triangle grid

Exercise 5.4.4 Show the following

1. A move does not change the Nim-sum of all vertices, when labeled as in triangular peg-solitaire.
2. Moves down will keep the total weight constant; all other moves increase the weight.
3. Compute the total weight of a board that starts at row 1 and looks like an inverted triangle, i.e., with i vertices in row i .
4. Place a single disk in row 1. Show that you cannot vacate the first four rows.

Solution: Summing over the diagonals:

$$\begin{aligned}
 \sum_{i=1}^{\infty} \frac{i}{2^i} &= \sum_{j=1}^{\infty} \sum_{i=j}^{\infty} \frac{1}{2^i} \\
 &= \sum_{j=1}^{\infty} \frac{1}{2^j} \sum_{i=0}^{\infty} \frac{1}{2^i} \\
 &= 2 \sum_{j=1}^{\infty} \frac{1}{2^j} \\
 &= 2.
 \end{aligned}$$

A single soldier in row 1, has weight $1/2$. Each move weakly increases the weight, so the total weight will stay at least $1/2$. If we are able to vacate the first n rows, the total weight will be less than $2 - \sum_{i=1}^n \frac{i}{2^i} = \frac{n+2}{2^n}$.

For $n = 4$, this number is $3/8$, which is less than $1/2$.

Thus, we are not able to vacate the bottom 4 rows.

What about the bottom three rows? For $n = 3$, this number is $5/8$, which is greater than $1/2$.

We now come to another puzzle invented by JOHN H. CONWAY and called CONWAY'S SOLDIERS. It is being played on the squares of an infinite chessboard, and we will call the soldiers just marbles. This board is divided by a horizontal line into an upper and a lower part. Initially, every square above the line is unoccupied, and the player can place any number of marbles below the line. A move consists of jumping with one marble across another marble horizontally or vertically onto an empty square, and removing the marble that was jumped over, exactly as in PEG SOLITAIRE.

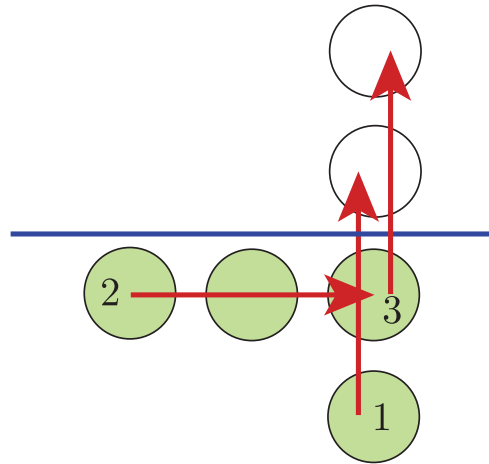


Figure 5.10 Moving to the second row

The puzzle asks to use a finite sequence of jumps to reach the fifth row above the line with at least one marble.

Exercise 5.4.5 Find a way to reach rows 1, 2, 3, and 4 with a finite number of jumps.

Solution: See Figure ?? for moving to the third row.

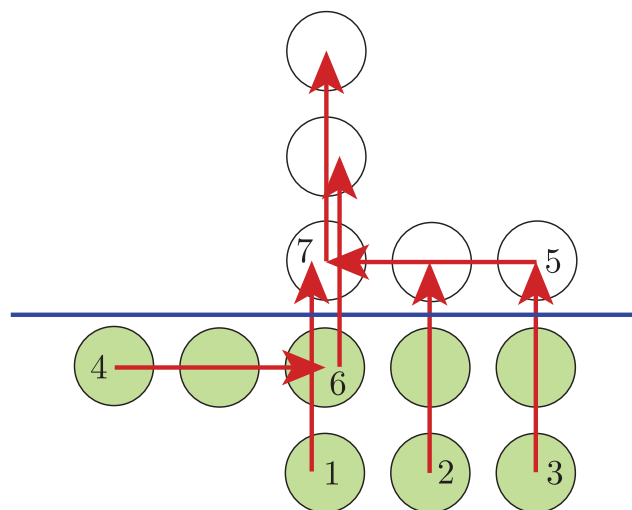


Figure 5.11 Moving to the third row

We will again put weights on the squares and measure the weights of sets of marbles. This time, the weights will be more complicated. Let ϕ be a positive real number to be determined. Put the weight $\phi^{|i|+|j|}$ on the square (i, j) .

Lemma 5.4.6 *If $\phi = (\sqrt{5} - 1)/2$, then no jump will decrease the total weight of a set of marbles.*

Proof. Note that $\phi^2 = 1 - \phi$. We make a case distinction.

1. If we jump from a square with weight ϕ^n away from $(0,0)$, we jump over a square with weight ϕ^{n+1} to a square with weight ϕ^{n+2} , thus increasing the total weight by $-2\phi^n < 0$.
2. If we jump from a square with weight ϕ^{n+2} towards $(0,0)$, we jump over a square with weight ϕ^{n+1} to a square with weight ϕ^n , thus changing the total weight by $\phi^n(1 - \phi - \phi^2) = 0$.
3. If we jump from a square with weight ϕ^{n+1} onto another square with weight ϕ^{n+1} , then we jump over a square with weight ϕ^n , thus changing the total weight by $\phi^n(\phi - \phi - 1) = -\phi^n < 0$.

Lemma 5.4.7 *The total weight W_1 of the region $j \leq -1$ below the horizontal line is $W_1 = (\phi + 1)^2$.*

Proof. We first compute the total weight of a quadrant $i, j \geq 1$. This is

$$\begin{aligned} \sum_{i,j=1}^{\infty} \frac{1}{2^{i+j}} &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \frac{1}{2^j} \\ &= \sum_{i=1}^{\infty} \phi^i \sum_{j=0}^{\infty} \frac{1}{2^j} \\ &= \frac{1}{1-\phi} \sum_{i=1}^{\infty} \phi^i \\ &= \frac{\phi}{(1-\phi)^2}. \end{aligned}$$

Thus, the total weight of the region below the line is

$$\frac{\phi + \phi^2}{(1-\phi)^2} = \frac{1}{(1-\phi)^2} = (\phi + 2)^2 \approx 6.85.$$

Corollary 5.4.8 *The total weight W_n of the region $j \leq -n$ below the horizontal line is $W_n = \phi^{n-1}(\phi + 2)^2$.*

Proof. The weight of the line at height $i + 1$ is equal to ϕ multiplied by the weight of the line at height i . □

We have the surprising

Theorem 5.4.9 *It is impossible to move a marble to the fifth row in finitely many moves.*

Proof. If this were possible, we also had a finite configuration in the region $j \leq -5$ with which we could reach the field $(0, 0)$. However, the weight of that configuration would be strictly less than $W_5 = \phi^4(\phi + 2)^2 = 1$. On the other hand, a configuration containing $(0, 0)$ has a weight of at least 1. □

5.5 INSTANT INSANITY

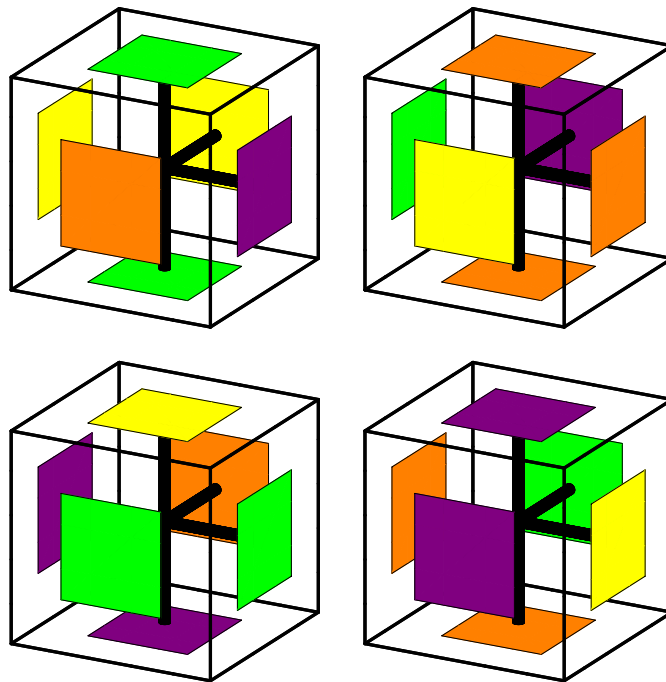


Figure 5.12 The cubes for INSTANT INSANITY

The puzzle INSTANT INSANITY consists of four cubes, colored with four different colors as shown in the nets in Figure 5.12.

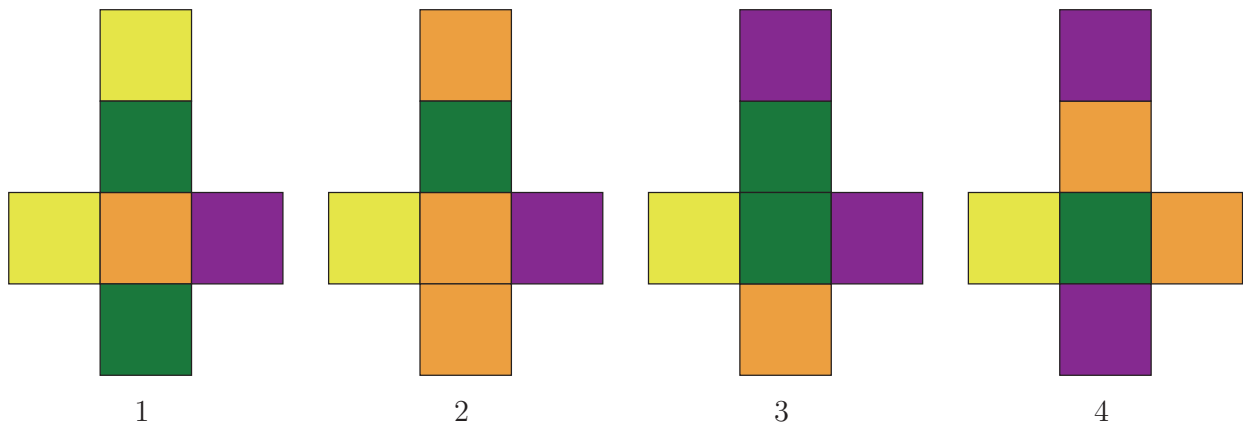


Figure 5.13 The nets for INSTANT INSANITY

The goal is to stack the four cubes together so that each of the four long faces of the $4 \times 1 \times 1$ tower shows all four colors.

According to JERRY SLOCUM, this puzzle goes back to 1900 when FREDERICK A. SCHOSOW marketed a version of it under the name KATZENJAMMER PUZZLE. It has since appeared in many variations. Its latest incarnation under the current name was discovered by FRANK ARMBRUSTER and has been marketed by PARKER BROTHERS since 1967.

We will solve the puzzle using graph theory. Each cube will get encoded in a graph. To do so, we use as vertices the colors (yellow, orange, purple, and green), and connect two colors by an edge if the two colors occur on opposite faces of the cube. Thus, we obtain for each colored cube a graph with four vertices (colors) and three edges (opposite faces).

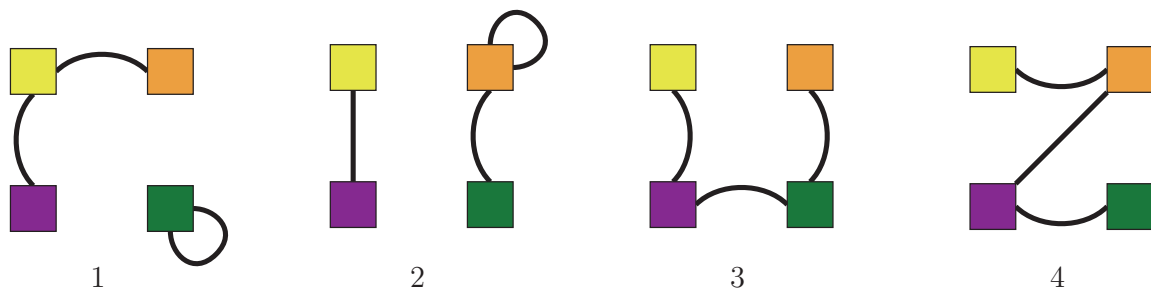


Figure 5.14 The opposite face graphs for each cube

Before we continue, let's see why this arbitrary-looking graph contains all the information about a cube that is necessary to play the puzzle: If the graph of a cube is given, we know all the colors of pairs of opposite faces for that cube. Using a blank cube, we can first color the front and back faces with a pair of these colors. It doesn't matter which color we pick for the front, because we could turn the cube over. Using a second edge, we then color the left and right faces with the two colors of the endpoints of the edge. Again, it doesn't matter which color we use for the left side, because there is a rotation fixing front and back that flips left and right. Finally, there are two possibilities remaining for coloring the top and bottom faces of the cube. These lead to truly differently colored cubes, but both are equivalent for solving the puzzle, as the top and bottom colors don't matter for the puzzle.

Coming back to solving the puzzle, we combine the four graphs we have created for each cube into a single graph with multiple edges. The edges are now labeled from 1 through 4 to indicate from which cube they come. How does this graph help us to solve the puzzle?

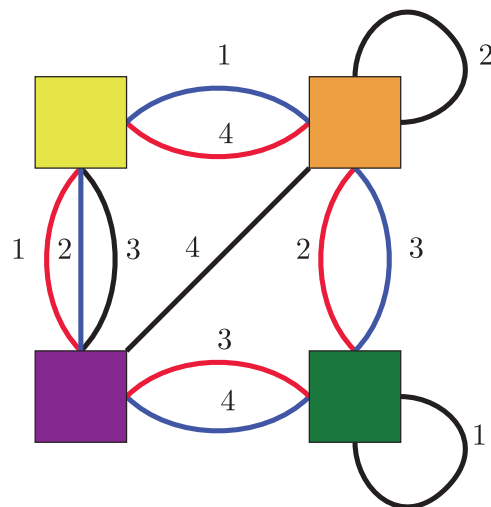


Figure 5.15 The combined puzzle graph

Suppose we have stacked the cubes together so that both the front and back side of the tower show all four difference colors. The four front and back sides of each cube represent each an edge in our graph, which we give a direction so that they always point to the color of the back edge. Let's mark these edges say blue. Thus we get four blue edges that begin at four different colors and end at four different colors. As there are just four colors, each vertex has an edge ending and an edge beginning there. This means that following the blue edges, we have found a doubly Hamiltonian circuit — a cycle or collection of cycles that pass through all four vertices (colors) of the graph, and uses each edges with each of the four labels (cubes). Vice versa, any such path can be used to stack the cubes so that front and back side of the tower show all four colors.

Next we aim to show that any such system of circuits must go once around the square marked by the four colors.

If this were not the case, it would decompose into several components of lengths $1 + 3$, $1 + 1 + 2$, or $2 + 2$. While in each case there are Hamiltonian circuits, none of them are doubly Hamiltonian.

Thus, we only need to look for Hamiltonian paths that go around the square, and can ignore the loops at orange and green, as well as the diagonal connection from purple to orange.

Consider the top and bottom edges, which both have an edge labeled 4. As each edge label can only occur once, one of these two edges must use an edge with a label different than 4. We make a case distinction: First, let's assume the bottom edge uses label 3. Necessarily, then, the right edge must use label 2, the left edge label 1, and the top edge label 4, and we get the Hamiltonian path marked red.

Now, let's assume that the top edge is using label 1. Thus, we have to use either label 2 or 3 for the left edge, and the other label 3 or two for the right edge. Either choice leaves us with label 4 for the bottom edge. One of the choices is the path marked blue; the other one has labels 2 and 3 exchanged in the left and right connections. Thus, there are only three different Hamiltonian paths.

To finally solve the puzzle completely, we will need to take care of the left and right faces of our tower. By associating a directed edge for each cube from every left face to every right face, we get a second doubly Hamiltonian path in our graph. This second path must be edge-disjoint from the first, as we cannot use a pair of opposite faces of the same cube twice.

As the second and third of our three Hamiltonian paths are not disjoint, one of the paths we seek must be the red path. This uses edge 2 on the right-hand side, so that the second Hamiltonian path can only be the blue one. They are indeed edge-disjoint, and thus solve the puzzle.

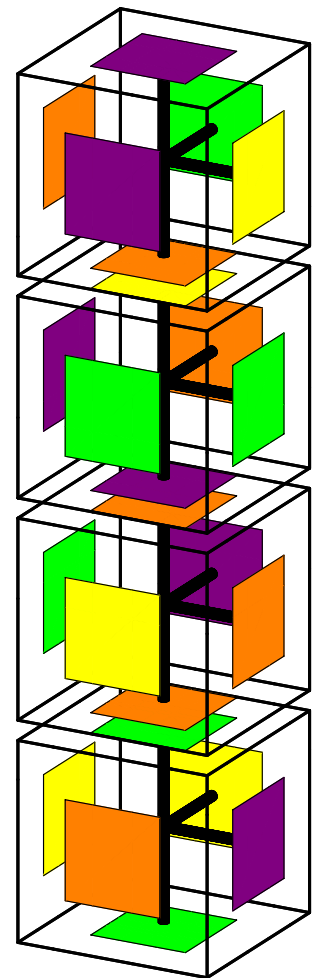


Figure 5.16 Insanity cured

6 Appendix

6.1 BIBLIOGRAPHY

- 1 J.F. Hurley, *Litton's Problematical Recreations*, Van Nostrand Reinhold Company, 1971.
- 2 E.R. Berlekamp, J.H. Conway and R.K. Guy, *Winning ways, for your mathematical plays*, Academic Press, London ; New York, 1982.
- 3 S. Loyd, *Sam Loyd's Cyclopedia of 5,000 Puzzles, Tricks and Conundrums with Answers.*, The Lamb Publishing Company, New York, 1914.
- 4 S.W. Golomb, *Polyominoes* (2nd ed.), Princeton University Press, 1996.
- 5 R. Honsberger, *Mathematical Gems I*, Mathematical Association of America, 1974.
- 6 J.H. Conway and J.C. Lagarias, "Tiling with polyominoes and combinatorial group theory", *J. Combin. Theory* 53(2), 183–208, 1990.
- 7 J. Kim and J. Propp, "A Pentagonal Number Theorem for Tribone Tilings", *The electronic journal of combinatorics* 30(3), 2023.
- 8 H. Rademacher and O. Toeplitz, *The Enjoyment of Mathematics: Selections from Mathematics for the Amateur*, Dover Publications, New York, 1990.

6.2 INDEX**a**

Archimedean tiling 85
 adjacency matrix 122
 adjacent 121
 axiom of choice 102

b

Banach-Tarski paradoxon 102
 Bubble Sort 154
 balanced polyomino 45
 bipartite 123
 bisector 118
 bridge 147

c

Conway, John H. 118
 chromatic number 146
 circuit 122, 136
 combinatorial game 3
 complete bipartite graph 123
 complete graph 122
 connected 124
 convex 139
 cycle representation 155

d

Dehn, Max 102
 degree 125
 diophantine equation 87
 directed graph 121
 disorder 154
 dissection 99
 dual map 146

e

Euler, Leonhard 136
 edge 121
 edge-to-edge 84
 even permutation 154
 exclusive or 13

f

Fibonacci numbers 76
 finite game 4
 fractal 15
 fundamental shape 81

g

Grundy number 26
 game tree 4
 generalized parallelogram 95
 generalized polygon 95
 golden ratio 76
 graph 121
 graph drawing 122
 group 14, 156

h

Hamiltonian cycle 50
 Hilbert, David 102
 half-square 55
 handshaking lemma 126
 hexagonal lattice 79

i

impartial 11
 isomorphism (graphs) 128

l

lattice 79, 96
 leaf 129
 length 124

m

Mock Turtle 40
 Morley's theorem 118
 map 140
 misère 15

n

Northcott's game 23

o

odd permutation 154
 option 3

p

Platonic Solids 139
 Platonic tiling 86
 Poker Nim 18
 parallel 95
 partition 125
 path 124

permutation 154
planar graph 140
polyform 42
polyhex 42, 69
polyiamond 42, 58
polyomino 42, 43
predecessor 121

r
Rims 23
ragged rectangle 53
regular map 142
rep-tile 52

s
Sierpinski tetrahedron 15
Silver Dollar Game 19
Sliding 18
NIM 11
NIM-sum 13
Swiss cross 103
simple graph 121
simple shape 95
snowflake 64
square lattice 79
star 123

strategy 4
subgraph 129
subtraction game 36
successor 121
sum (of games) 27

t
Tree Pruning 130
tiling condition 97
translation 78
translation invariant 78
tree 129
triangle 138
trisector 118

u
uniform tiling 85

v
valency 125
vertex 121
vertex transitive 85

w
winning position 4
winning strategy 4